# 语气与声调加工的机制及其受音乐障碍的影响


李嘉琦*

(北京师范大学认知神经科学与学习国家重点实验室，北京 100875)




## 摘 要


背景：在声调语言里，语气和声调都依赖于音高线索，语气能帮助听话者理解意图，声调能够指示字词的意思，这两者也在句子中有所交互。然而在动态过程中这两者被大脑加工的先后层级还尚不清楚。此外，音乐与语言是两个联系紧密的领域，音乐上的障碍如何影响人们对言语中语气和声调的感知，在成人和儿童中是否有类似的影响，也仍待研究。

目的：本文主要研究汉语母语者进行语气与声调加工的机制及其与音乐障碍之间的关系，分为两个研究问题（1）语气和声调加工的时序层级；（2）音乐加工障碍对语气和声调识别的影响。

方法：本研究的被试群体包括 11 名成人音乐障碍被试，13 名成人控制组被试，21 名儿童音乐障碍被试，23 名儿童控制组被试。实验材料是找普通话标准的录音者录的朗读汉语句子的音频。在第一个实验中，对 13 名音乐能力正常的成人播放表达不同语气和声调的汉语句子，让他们判断语气/声调，并记录下他们的行为表现和脑电。实验二关注音乐能力对语气、声调判断能力的影响，对音乐能力障碍的 11 名成人被试，以及不同音乐能力的儿童被试播放同样包含了语气和声调信息的句子，记录他们在语气和声调识别上的正确率。

结果：实验一发现在语气任务中，大脑对语气信息的反应在很早的时间窗就出现了，对声调信息的反应后一步出现。而在声调任务中，脑电上最早出现的效应则是语气和声调的交互作用，揭示了语气信息无论在语气还是声调判断上的重要参与。实验二发现音乐障碍无论在成人组还是儿童组都主要影响语气加工，而不是声调判断能力。此外，儿童的数据显示，语气加工的能力要早于声调发展成熟，这个结果可能也与语气对人类社会沟通的重要性有关。



**关键词** 语气 声调 汉语 先天性失乐症 音乐能力



* 李嘉琦 E-mail: jiaqi.li@mail.bnu.edu.cn

# THE IDENTIFICATION OF INTONATION AND TONE IN MANDARIN AND THE INFLUENCE OF PITCH DISORDER ON IT

## ABSTRACT

Background: Intonation and lexical tone both rely on F0 in tonal languages. While the former one can indicate the speaker's intention (e.g., raise a question or state a fact), the latter helps determine the meaning of a word. These two interact with each other in sentence, but the access order of them by the brain is unknown. Besides, as a closely related domain with language, music might have some influence on the processing of intonation and tones. Whether the music disorder affects adults' and children's intonation and tone identification ability still remains unclear.

Purpose: The current study aimed at investigating: (1) the access order of intonation and lexical tones in Mandarin Chinese; (2) whether people who are impaired in pitch perception and who have normal music ability perform differently in intonation identification and tone identification tasks.

Method: Eleven adults and twenty-one children who have pitch disorder, thirteen adult controls and twenty-three children controls participated in this study. They were required to listen to Mandarin sentences containing both intonation and lexical tone in the same syllable, and judge the intonation (in the intonation task) and tone (in the tone task). In experiment 1, thirteen adults with normal music ability participated, and their behavioral reactions and mean amplitudes in ERP were recorded and analyzed. In experiment 2, the intonation and tone identification of adult and children amusics (as well as children controls) were investigated. Their behavioral performance was recorded and compared in data analysis.

Results: In experiment 1, according to the ERP results, it was found that in the intonation task, the participants responded to intonation first and then the lexical tone. In the tone task, participants were sensitive to both intonation and tone in an early interaction beginning from 100ms. Participants were alert to intonation and would process it no matter in what tasks, which indicated the important role of intonation in human communication. In experiment 2, overall, music disorder impaired people in the identification of intonation, rather than lexical tones, in both adults and children. Besides, the perception of lexical tones has not been fully developed in 9-year-old children, while their ability to identify intonation was mature, which indicated that intonation identification ability might be mature earlier than tone identification in human beings.

**KEY WORDS：** Intonation, lexical tone, Mandarin, amusia, music ability

# 目 录

# 第 1 章 引言

## 1.1 语气和声调加工的时序层级

音高（pitch）是语言中非常重要的一个特征。在句法层面，它与语调有关，能够指示语调短语边界和语法边界[1, 2]；在词法层面，它与声调有关，在声调语言中可以决定一个词汇的意思[3]。语气和声调都依靠 F0 线索，这二者不可避免地在言语加工的过程中有所交互[普通话研究见 4, 5, 6, 粤语研究见 7]。前人研究非常详尽地探究了声调对于语调的影响，并发现了在下降声调（例如汉语中的四声）和上升的疑问语气之间的冲突[5, 6, 8-11]。但是，无论是对语气还是声调的加工都是随着时间逐渐展开的，在对语气和声调判断以及这两者的交互的研究中，这一时间动态性需要被关注。然而迄今，在一个表达中语气和声调到底是何时被加工的、这两者的加工顺序，都尚不清楚。因此，在这个研究的第一个实验中，研究者详细探索了语气和声调加工的时序层级以更好地了解言语韵律加工的动态过程。

为了探究语气和声调在声调语言中的加工顺序，在这里本文研究者提出了两个假设：任务相关假设和语气优先假设。（1）任务相关假设：人们在很早之前就发现语言的实时加工是依赖于任务内容的[12]。采用相同的语言信号（辅音-元音-辅音的音节对子），语音辨别任务会导致左侧布洛卡区活动的增强，但在音高辨别任务中，则是右侧的前额皮质被激活了[12]。如果语气和声调的时序层级在本质上也是依赖于任务的话，那么在声调任务中，与声调相关的效应就会率先被观察到，与之相反，在语气任务中，语气效应则是具有时间优先级的。（2）语气优先假设：在言语加工中，语气可能比声调有更高的时序层级。有以下四个理由支持该假设。首先，与声调相比，韵律线索可能具有更深的进化的根源，而且可以脱离于人类语言独立存在。对于禾雀，韵律辨别是一种主要的能力。研究发现在日语中，它们能注意到韵律中的差异，而不是句子内容上的差异，即使这两者它们都有辨别能力[13]。Spierings 和 ten Cate 发现斑胸草雀对跟人类言语中相同的一系列韵律线索（音高、长度、振幅）敏感。此外，音高线索比其他的线索得到了更突出的处理，与人类感知言语韵律时的方式很像[14]。此外，一个交叉抚养实验[15]发现，斑胸草雀鸟鸣声韵律的时间结构是先天的。支持语气优先假设的第二个原因是，在很早期就发展了的韵律技能对新生儿的言语学习起着关键作用。有研究发现，五个月大的婴儿可以对赞成/不赞成[16]以及积极和消极的表达[17]有不同的回应，而且婴幼儿在六个月到九个月大的时候就发展了对韵律特征的偏好[18]，婴幼儿可能依靠韵律线索（例如重音模式）来从流畅的言语中切分出词语[19]。第三，语气是不同语言普遍具有的[20]表达特征，它不

仅传达了言语信息，也包括了人类交流的副语言信息，例如说话者的情绪、态度和目的[20-23]。人们甚至可以从不熟悉的西方或者非西方语言中提取出情绪韵律[24]。这样的社会性的信息对人类的社会交流和自身生存发展具有重要的作用[例如 25]，因此它们可能比起词汇信息在大脑中的加工上有更高的优先级。第四，前人研究发现，人们对于声调具有范畴知觉[26, 27]，但是对于语气却没有[28]，考虑到我们的大脑会将有限的资源优先给与最重要的任务，可能微小的语气上的改变都对人类很重要。在 Liu 和 Rodriguez 的研究中，他们在英语母语和汉语母语的群体中考察了英语语气（疑问/陈述）的范畴知觉，比起英语母语的被试，汉语母语被试表现出了更陡的语气识别函数和不同的语气边界。然而，分析得到的辨别函数并不是典型的范畴知觉的函数，两组都没有出现（或只出现非常小的）辨别峰值（discrimination peak）[28]。

考虑到语气与声调相比具有的以上优势与重要的社会意义，人们可能对于语气比对声调更敏感。但是之前的大部分研究都是关注在语气和声调是怎么交互的，缺少对它们在动态的加工过程中的时序层级的比较。研究者广泛认为声调能够影响语气的识别，当末尾音节是四声的时候比二声时更好识别出疑问语气，而对于陈述语气的识别没有发现声调影响的效应[5, 6]。四声下降的声调与疑问语气的上升轮廓冲突，这不仅使得疑问语气更容易被识别出来，也加强了四声刺激的疑问和陈述语气之间的差异对比[关于疑问和陈述语气在四声上而不是二声上出现的 MMN 请参考 8, 9, P300 成分请参考 10, P600 成分请参考 11]。至于语气对声调的影响，之前的研究认为在汉语中，语气不会影响声调的轮廓[29]，对声调的感知也几乎没有影响[4]。汉语的婴儿指向型言语（具有夸张的语气轮廓）也并没有扭曲声调[30]。考虑到下降的声调对于疑问语气识别的强大干扰作用，Yuan 从一个局部视角将语气视为依赖于声调的，总结说声调识别可能领先于语气的识别，或者至少同时被加工[6]。然而，一个广东话研究发现疑问语气会使得大部分低声调被错误识别为高声调[31]，这表明语气导致的 F0 轮廓上的变化可能对声调的识别是有影响的。

探究语气和声调加工的时序层级对于更清楚地分析声调和语气的识别哪一个在声调语言中有优先地位有着重要作用。我们需要实时了解整个动态过程的较高时间分辨率水平上的语气与声调加工细微差异。事件相关电位（ERP）是一个很好的考察这个过程的研究手段，因为它具有高时间分辨率，可以探测到听觉线索可能对于大脑活动产生的影响。因此，研究者可以发现语气和声调在听觉加工时是什么时候参与进来的，并且比较两者的时间先后。为了给语气一个公平的比较机会，本研究也设置了一个声调识别的任务，而不是只有语气识别任务，这样语气线索对声调识别的影响也可以被考察到。Liu 和她的同事在研究中也设置了语气和声调两种任务[10]，但她们的 ERP 数据分析混合了两个任务中得到的脑电数据，并没有分别进行分析，这样使得研究者难以比较语气和声调加工各自的情况，也很难确定语气和声调何者具有更强的敏感性，更容易被汉语母语的被试所感知到。因此，为了比较这两者的效果和探索自上而下的加工情况，语气和声调识

别的任务都应该被研究和分析。

## 1.2 音乐加工障碍者的声调和语气加工

音乐在人们的生活中扮演着重要作用[32-34]。很多研究都表明，音乐能对其他领域的技能和认知能力产生影响，例如语言[35-38]、数学[39]和空间推理[40-44]。然而，Kalmus 等人[45]的研究指出，英国接近 4%人口都有一种影响音乐音高加工的障碍，这种障碍被称为先天性失乐症（congenital amusia[46]）。与此类似，在中国，这种音高加工障碍在人群中的比例也达到了 3.4%[47]。先天性失乐症是一种在音乐感知和产出上存在困难的障碍[48, 49]，以音高加工困难为主要特征[50]。失乐症患者在感知、记忆和欣赏音乐方面的能力存在一定程度的缺陷[46]，而且这种音乐方面的障碍不能归因于听力、智力和音乐经验的差异[49, 51]。 鉴定失乐症目前最广泛使用的测验是 Peretz 等人编制的“蒙特利尔失乐症成套测验”（The Montreal Battery of Evaluation of Amusia，简称为 MBEA）[52]。该测验包含 6 个分测验：与音高有关的 3 个分测验（分别考察音阶、异轮廓、同轮廓）；与节奏有关的 2 个分测验（节奏识别测试、节拍识别测试）；以及与记忆有关的 1 个分测验（旋律再认测试）。

同样作为人类生活中必不可少的基础部分，音乐和语言在演化和认知加工上有着紧密联系[53]。它们在很多方面都有相似之处，具有相似的声学物理特性和句法结构[33, 34, 54, 55]，而且部分共享了相同的认知神经机制[53]。音高加工对于音乐中的旋律和语言中的韵律来讲都同样是很基本的[56]。有研究表明，音高方面的经验可以在这两个领域中相互迁移[57-59]。不过目前，对于音乐和语言领域中的音高加工在多大程度上是重合的，研究者还没有得到一致的定论。因此，对音高感知的研究或许能够帮助探索音乐与语言的相似和不同之处。研究失乐症患者在语言相关的任务中的表现，能够给研究者一个机会探索音高加工在这两个领域中的异同。

前人的研究已经发现，在声调感知、语音加工、语调和情绪韵律加工和句子理解上，失乐症患者都表现较差的能力[47, 60-66]。Nan 等人发现将近三分之一被施测的汉语母语失乐症被试在声调的识别和区分中有障碍[47]，Nguyen 等人的研究发现法语母语的失乐症在辨别汉语声调上也差于控制组[67]。

在语气的识别和区分上，Ayotte 等人[51]在蒙特利尔的研究发现，在识别言语中的语气（即识别陈述语气和疑问语气）和语气辨别（判断语气是否相同）的任务中，失乐症群体并没有与正常被试的表现有显著差异。不过，当句子中的所有言语信息都被移出后，失乐症表现出能力上的下降。与之类似的是，Patel 等人[66]的研究也发现，失乐症根据语气区分句子的能力远远好于区分表示语气轮廓的离散的音高类似物（discrete-pitch analogs），而且他们在辨别滑顺流畅的音高类似物（gliding-pitch analogs）上也表现出了

缺陷。同时，也有研究揭示了失乐症在言语语气判断上的缺陷。例如，研究者发现英国的失乐症在陈述句与疑问句的区分、辨认和模仿上存在问题，这里的陈述句与疑问句最主要的差别就在最后一个字的音高方向[64]。在此基础上，Hutchins 等研究者探索了失乐症在言语任务和非言语任务上的表现的这种差异是否跟音高变化的范围有关系，他们认为，这种音高变化在言语中是比较粗糙的，而在音乐中相对更精细[68]。当失乐症被试们将句子归为陈述句或者疑问句时，研究者通过修改最后一个音的 F0 值（从第一个水平到第十一个水平），发现失乐症群体用了和正常被试相同的对刺激进行归类的标准，虽然失乐症的正确率和内部一致性更低。失乐症被试似乎在两个端点之间的精细变化上有问题，说明音高感知障碍影响了他们在音高轮廓上的能力，因此证实了作者的假设，失乐症在音乐和语言上的感知能力不同主要是因为语言中的音高差异并不像音乐中那么精细。

以上的发现都提示了我们探究失乐症的领域特异性的重要性[53]，然而，与此同时，关于声调语言背景中的失乐症的语气研究相对来说就比较少了。值得一提的是，在声调语言中，音高变化对单个文字而言是一种识别特征，因此研究声调语言母语者的失乐症患者能帮助探索音高加工缺陷是否能够迁移到语言领域[68]。接下来本文作者将会介绍这些数量较少的关于声调语言国家的失乐症患者在语气判断上的表现的研究。

Jiang 等研究者 2010 年在中国做了两个实验，让失乐症被试完成对旋律轮廓以及对言语语气的区分和识别任务。在语气任务中，实验材料是两个音节构成的普通话的动宾短语，被试需要区分出陈述语气和疑问语气的刺激。结果发现，相对于正常被试，失乐症被试不仅在辨别旋律轮廓上有缺陷，而且也在言语和非言语类似物的语气上出现了问题[69]。

有研究者发现，失乐症是在主要依靠音高的灵敏度的任务上的表现有所缺陷，例如区别从陈述句和疑问句上提取出的平滑的音调序列，而在需要利用多个声学线索得到意义的任务上没有明显问题，例如词语识别任务、自然言语中的陈述-疑问句识别与区分的任务[62]。与上一个研究[69]相似，在这个研究中，Liu 等人所采用的实验材料依然是一对对单字一样，但语气上不同的词语，然而，他们用的词语从三个音节到七个音节不等，陈述语气的词语和相对的疑问语气词不仅在 F0 上有差异，而且音长（duration）和声强（intensity）上也不同。配对 t 检验显示，陈述语气的刺激比疑问语气的刺激具有显著更低的平均 F0、更低的平均声强、更宽的音高范围，这既体现在整条词语刺激上，也符合最后一个音节的情况。而且，结尾是陈述语气的音节比疑问语气的音节有显著更短的音长。而在之前 Jiang 等研究者 2010 年的研究[69]中，他们用的词语对子是两音节的，而且只在第二个音节的 F0 曲线上有差异。这两个研究在使用的刺激上的差异或许可以解释为何他们得到的实验结果不一样（即为何在 Jiang 等人的研究中，汉语失乐症被试并没有与控制组在语气任务上表现得一样好）。这提示我们，或许其他声学线索，例如音长和声

强，能够帮助汉语母语的失乐症群体进行言语相关的加工。

以上这两个关于中国的失乐症群体的语气加工的研究都是用短语当作测试材料，与它们不同，Jiang 等研究者[60] 2012 年的研究中使用了对话而非单字或短语来作为实验刺激。这个研究探索了失乐症对日常言语理解中的韵律和句法不匹配的大脑反应。被试需要判断包括了疑问/回答的对话中的语义可接受性。在表示回答的句子中，有“是”或“不是”，一个解释原因的从句，以及一个意义上相关联的表示陈述语气或疑问的从句。最后一个音节要么被说成了疑问语气，要么是陈述语气，使得整句话的语气可能合适，也可能不合适。与控制组相比，失乐症被试在这个任务上表现得更差。同时，失乐症被试在脑电的 N100 和 P600 成分上并没有表现出语义不合适条件和合适条件之间的显著差异，而控制组对于不合适的刺激表现出了更大的 P600 和更小的 N100 成分。

近年来的这些对于失乐症在语气判断上的表现的研究对探索失乐症的语言能力以及音乐与语言关系都做出了很大的贡献，不过在句子语调的加工过程中，除了以上研究中广泛使用的语气线索音高以外，还有一些其他的线索发挥着作用，例如语气中的声强，它也可能为失乐症的言语理解提供了很大的补偿作用，也应该被考虑进研究中。其次，由于上文总结的语气和声调的紧密联系，在考察失乐症的语气和声调判断时，也应该考虑到声调和语气的交互作用，探究失乐症在识别受到语气干扰的声调（例如疑问语气时的四声）和受到声调影响的语气时，是否存在和正常被试群体之间的差异。此外，关于儿童音乐障碍群体的语言加工能力的研究较为缺少，需要有研究来考察在以声调语言为母语的儿童里音乐障碍对语气和声调识别的影响。

# 第 2 章 研究问题



## 2.1 语气与声调加工的时间进程次序

声调和语气在实时言语感知中彼此交互。先前的研究没有探索声调和语气在同一音节中被加工的时序特征，在本研究中，设计了两个任务（语气识别任务和声调识别任务）来探究——我们的大脑加工这些信息时候的次序。

因此，在这个研究中，ERP 脑电效应出现的时间是一个重要的因变量，自变量包括了听觉刺激的语气和声调信息。研究者需要采用可以同时表征语气信息和声调信息的刺激材料，汉语普通话符合这一需求，因为在汉语中，声调和语气可以被同一个末尾音节承载。事实上，前人研究已经表明在汉语中语气感知更多地依靠末尾音节的 F0 特征而不是整个句子的音高水平[70]。F0 是声调的主要线索，而对于语气来说，研究者认为疑问句比陈述句有更高的 F0 和声强[5]，但声强对语气识别的贡献有多大仍在讨论中[71]。为了彻底研究语气的听觉线索是什么时候、怎么样在加工过程中出现的，本研究操控了音高和声强两个线索。研究者将语气的变化分解为 F0 和声强的变化，控制了音长线索和句子内容。因此，总的来说有语气-音高（疑问音高和陈述音高）、语气-声强（疑问声强和陈述声强）以及声调（二声和四声）三个自变量。采用二声和四声是因为这两个声调的特殊性。正如 Jiang 和她的同事在研究中所总结的，在汉语中，对于疑问语气，除四声以外的其他声调都是向上的，只有四声的音高轮廓是下降的；对于陈述语气，二声是向上的趋势，而其他声调都是向下的[69]。虽然总体上疑问语气会有一个上升的音高走向，陈述语气会对应着下降的轮廓，但汉语的二声和四声这两个声调还保持着自己的特征，这点便于研究者探索声调和语气各自是怎么加工的以及它们之间的互相影响。本研究实验一的实验假设是：（1）无论是在汉语的语气识别任务还是声调识别的任务中，语气都是先被加工的。（2）声调的听觉线索会对语气识别有影响，同时语气也会影响声调识别。

## 2.2 音乐障碍对语气与声调加工的影响

音乐与语言两个领域联系紧密。从以前的研究可以看到音乐障碍可能会影响语气的判断，但具体对语气中音高、声强因素是怎么影响的还不够了解。本研究在探索句子语气加工时，除了使用了前人研究中广泛使用的 F0 线索以外，还使用了其他的语言线索，例如声强，很好地控制了参与语气判断的各个线索，包括音高、声强、音长，其中声强很可能为失乐症的言语理解提供了很大的补偿作用，这一设计能帮助我们更好地了解音乐障碍对失乐症患者语言感知的影响。

通过本实验中刺激材料的正交设计（陈述音高-陈述声强、疑问音高-疑问声强、陈述音高-疑问声强、疑问音高-陈述声强），可以探究失乐症和控制组被试分别在什么样的材料中能判断得好，什么类型的材料会不利于语气识别；两组如果存在语气识别能力的差异，是在所有材料中存在还是某一类刺激类型中存在的。此外，通过设置声调任务，本实验也探索了失乐症和控制组在声调判断上是否存在差异，这两组的声调判断是否受到语气因素的影响。

同时，以前大部分探讨音乐能力对语气、声调判断的影响的研究都是基于成人被试的，本实验还招募了小学儿童被试，以便探察成年人组中音乐障碍带来的效应是否在儿童群体中存在。通过对比儿童中不同音乐能力的群体在语气和声调识别任务上的表现，可以考察音乐加工障碍在儿童的语言加工上的影响；通过对比同样具有正常音乐能力的儿童和成人在两个任务上的情况，可以分析人类的声调、语气加工的能力的发展和成熟规律。

# 第 3 章 实验一：语气和声调加工的时序层级

## 3.1 实验方法

### 3.1.1 被试样本

研究者通过在校园论坛发被试招募帖子，招募到 13 个音乐能力正常、汉语母语、且没有接受过超过两年的正规音乐训练的被试（5 男 8 女）。他们都没有报告有任何神经或者心理疾病的相关病史，而且在 250 至 8000 赫兹（Hz）的最小听力阈值均低于 20 分贝（dB），具有良好听力，问卷测得他们都是右利手[72]。在蒙特利尔失乐症成套测验（MBEA）[52]中，他们都处于正常分数段（高于鉴定标准 71.7%）。

### 3.1.2 刺激材料

本研究用了 39 对汉语普通话中的单字，每对单字中的两个字具有相同的元音和辅音，只在声调上不同（一个是二声，一个是四声），所以一共组成了 78 个单字刺激，它们都是真字。实验材料由一句话的主干“这个字念”加上述单字作为句子末尾字组成（例如“这个字念‘替’”）。这句话是中性语义的。研究者通过一名年轻的汉语普通话为母语的女性朗读这 78 句话的录音，获得原始的语音刺激材料。朗读者在读每句话时会读两遍，既会读出这些句子的陈述语气，也会再读一次句子表达疑问语气时的版本。因此一共读了 156 个句子。随后，研究者将每个句子的末尾字剪切下来，得到了 156 个单字的音频材料。对于每个单字，都有两种声调，每种声调有两个语气，也就是一共四种刺激材料。研究者用 Praat 音频处理软件[73]将每个单字的四种刺激的音频的时长（duration）进行了统一，具体方法是先提取这四个音频的时长，得到其平均值，再将平均时长赋予这四个音频。这样一来，每个单字内部的不同条件具有同样的时长，不同字的时长不一样，既控制了时长这一语言因素，又保持了一定的生态效度。研究者再选取了一个表达陈述语气的句子主干“这个字念”（时长为 989ms），跟这些单字（时长在 405 到 584ms 之间）重新组合，得到 156 个句子。这样一来，所有句子的主干都是相同的，只在最后一个字上不同。

本文研究者招募了 11 名汉语母语的大学生被试来对这些句子进行判断，每个被试要做两种任务，在语气判断任务中，被试要判断这个句子表达了什么语气（陈述句按 F 键，疑问句按 J 键），在声调判断任务中，被试要判断这个句子的末尾字是几声（二声按 2，四声按 4）。任务的呈现顺序随机。对于任何语气或声调判断正确率在 0.72 以下的刺激，相应的字（包括它的四种刺激类型）都会被剔除出实验材料。总共有 28 个句子被移

除，剩余 128 个句子保留了下来（32 个字对*2 声调*2 语气）。语气判断的平均正确率为 0.93 (最小值：0.73；最大值：1.00；标准差：0.09) ，声调判断的平均正确率为 0.98 (最小值：0.82；最大值：1.00；标准差：0.05)。此外，二声和四声的单字材料，在词频上不存在显著差异(t(57) =1.617, $p$=0.111)。通过进一步的声学分析，发现对于每个单字而言，它的疑问刺激和陈述刺激在 F0 level(t (63)=25.046, $p$<0.001)和声强(t(63)=6.783, $p$<0.001)上有显著的差异。平均 F0 level、声强和时长都在表 1 中显示了。图 1 是一个示例刺激（“fei”）的 F0 轮廓图。

除此之外，研究者也通过计算 0% 与 100%时间点的半音程差异（100%时间点-0%时间点）得到了 F0 interval。在 F0 interval 和 F0 level 上，对所有末尾单字刺激的方差分析发现了显著的语气主效应（对于 F0 level：F(1,31)=592.736, $p$<0.001, $\eta_p^2$=0.950；对于 F0 interval：F(1,31)=82.936, $p$<0.001, $\eta_p^2$=0.728）和声调主效应（对于 F0 level：F(1,31)=539.493, $p$<0.001, $\eta_p^2$=0.946；对于 F0 interval：F(1,31)=431.720, $p$<0.001, $\eta_p^2$=0.933），见图 2。然而，无论对于二声还是四声，陈述都比疑问刺激有更低的 F0 level ($p$s<0.001)和 F0 interval ($p$s<0.001)。无论对于疑问还是陈述语气，二声都比四声刺激有更低的 F0 level ($p$s<0.001) 和更高的 F0 interval ($p$s<0.001)，语气和声调之间并没有有意义的交互作用（图 3）。

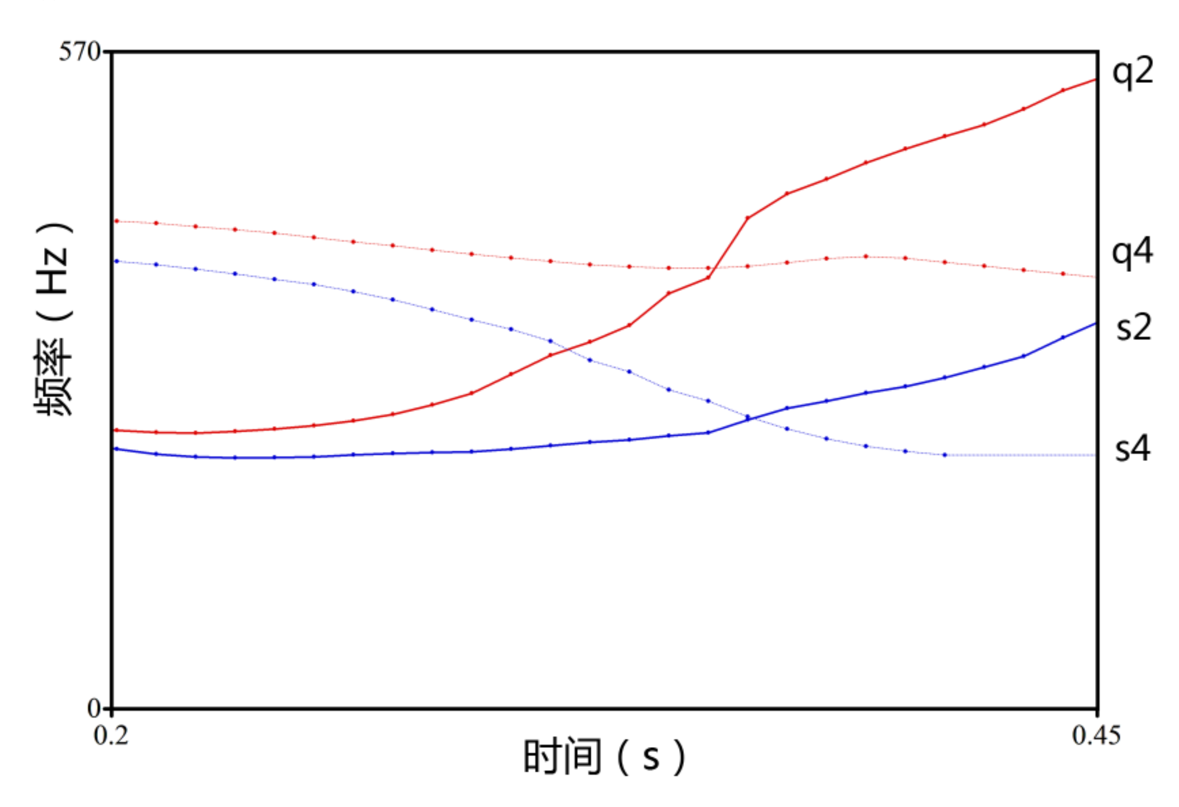



图 1 示例刺激（“fei”）的基频轮廓图（F0: 0–570 Hz）。图中呈现了两个声调（二声和四声）和两种语气（疑问和陈述语气）。

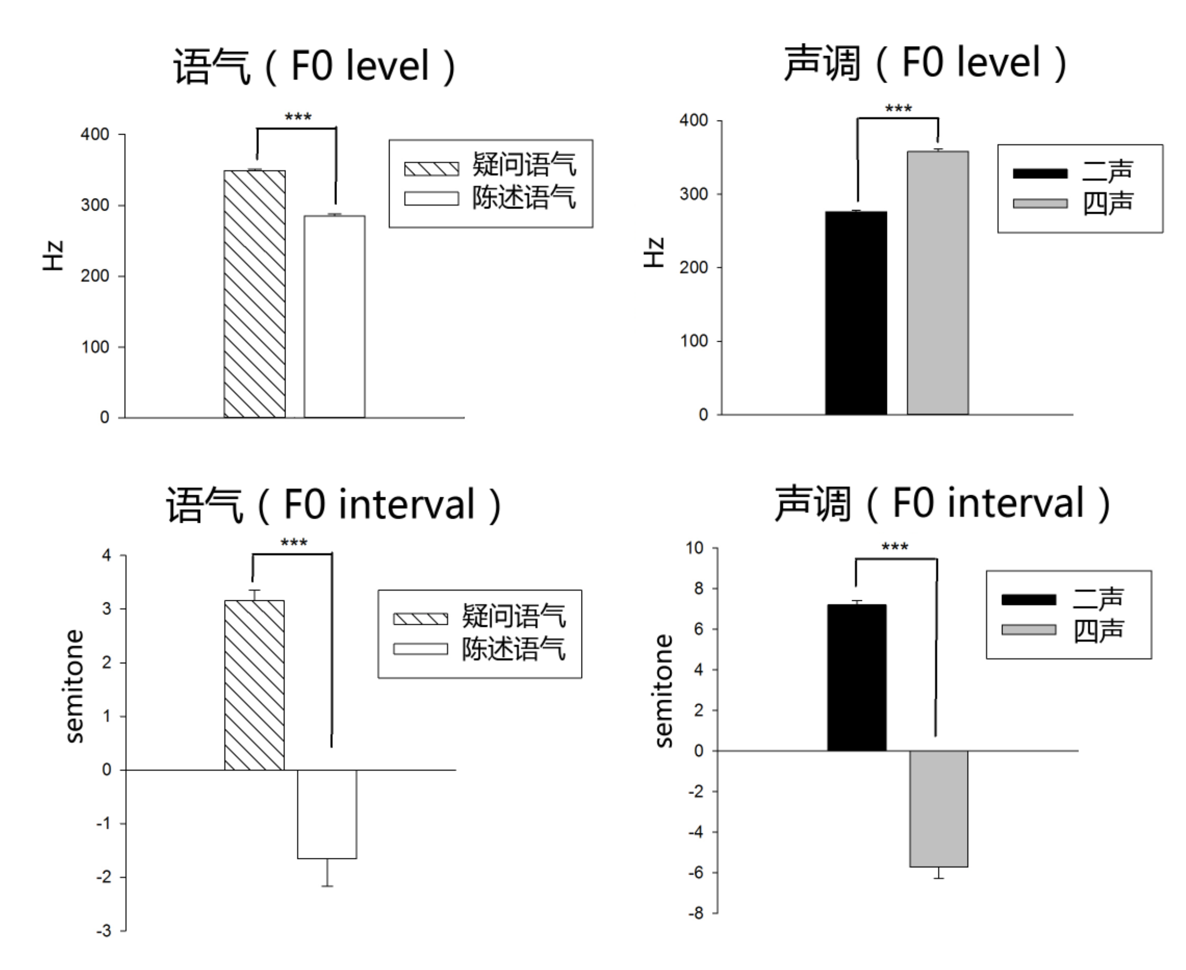



图 2 语气和声调在 F0 level（上）和 F0 interval （下）的主效应（*n.s.*表示 $p>0.05$，*表示 $p<.05$，**表示 $p<.01$，***表示 $p<.001$；误差线表示标准误）。

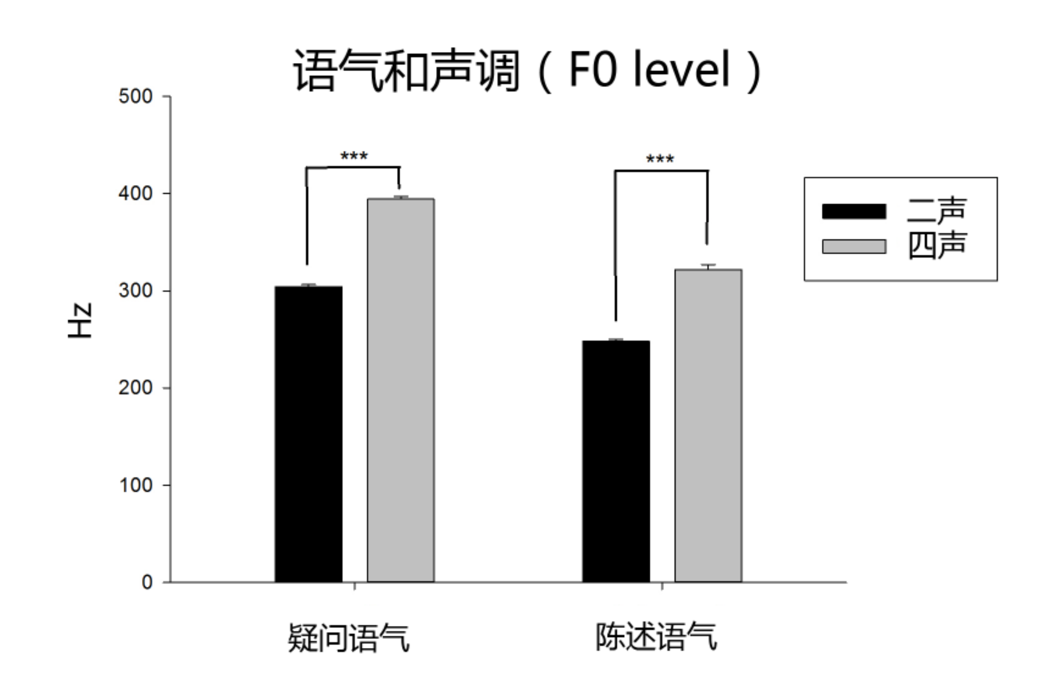


图 3 语气和声调之间不存在有意义的交互效应（*n.s.*表示 $p$>0.05，*表示 $p$ < .05，**表示 $p$ < .01，***表示 $p$ < .001；误差线表示标准误）。

为了探究音高和声强在语气加工中的作用，研究者用 Praat 音频处理软件[73]将每个末尾单字的陈述语气材料与疑问语气材料的音高（pitch）进行互换，对包括声强（intensity）在内的其他语言线索不进行改动。研究者因此得到了 128 个新的单字音频（32 字对*2 声调*2 操作后的语气），这些字既不是单纯的陈述语气也不是疑问语气。与之前相同的句子主干加在了这些新字的前面，组成了 128 个新的句子。这些不自然句子加上之前的 128 个自然句子，总共有 256 个句子作为实验材料。对于每个单字的每个声调而言，都有四种语调刺激（陈述语气材料、疑问语气材料、音高变为疑问语气音高之后的陈述语气材料、音高变为陈述语气音高之后的疑问语气材料）。因此，在本研究中，总共有八种实验刺激类型，包括四种自然刺激（q2:二声的疑问刺激, q4：四声的疑问刺激, s2：二声的陈述刺激, s4：四声的陈述刺激），以及四种不自然刺激（s2(q)：音高变为陈述语气音高之后的二声疑问语气刺激, s4(q)：音高变为陈述语气音高之后的四声疑问语气刺激, q2(s)：音高变成疑问语气音高之后的二声陈述语气刺激, q4(s)：音高变成了疑问语气音高之后的四声陈述语气刺激）。表 1 是这八种刺激末尾字的声学参数。

表 1 八种刺激末尾字的声学参数(2 声调*2 语气*2 操作后的语气；SD 表示标准差)。

| 刺激类型 | F0 level (Hz) | | F0 interval (semitone) | | 声强 (dB) | | 音长 (s) | | 数量 |
|---|---|---|---|---|---|---|---|---|---|
| q2（SD） | 304.33 | (12.54) | 8.43 | (2.12) | 73.71 | (1.97) | 0.49 | (0.04) | 32 |
| q4（SD） | 394.33 | (17.32) | -2.11 | (1.49) | 76.89 | (2.40) | 0.49 | (0.04) | 32 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| s2（SD） | 248.30 | (10.32) | 5.99 | (1.33) | 72.47 | (2.05) | 0.49 | (0.04) | 32 |
| s4（SD） | 322.00 | (28.76) | -9.30 | (5.93) | 75.40 | (2.62) | 0.49 | (0.04) | 32 |
| s2(q) （SD） | 247.54 | (10.83) | 6.02 | (1.33) | 73.71 | (1.97) | 0.49 | (0.04) | 32 |
| s4(q) （SD） | 324.56 | (23.37) | -10.01 | (3.91) | 76.89 | (2.40) | 0.49 | (0.04) | 32 |
| q2(s) （SD） | 302.27 | (15.65) | 8.75 | (2.89) | 72.47 | (2.05) | 0.49 | (0.04) | 32 |
| q4(s) （SD） | 393.30 | (17.71) | -2.60 | (2.45) | 75.40 | (2.62) | 0.49 | (0.04) | 32 |

### 3.1.3 刺激呈现

实验中，被试先看到指导语页面，然后屏幕中心出现注视点（250ms），接着播放音频材料的句子，其中句子主干“这个字念”的时长为 989ms，句子末尾单字的时长在 405 到 584ms 之间，脑电的 trigger 加在末尾字出现前。从末尾字开始播放到反应阶段之间有 1200ms 的时间，当末尾字播放完，也不会翻页，等 1200ms 结束后才会进入反应阶段，屏幕上出现一个“请按键”，反应阶段开始。被试需要在听到刺激后尽可能又快又准确地在键盘上按下相应的反应按钮，反应阶段的最大时长为 3000ms。

### 3.1.4 实验流程

本实验分为两个任务，语调判断任务和声调判断任务，两个任务的顺序随机。两个任务所用的刺激材料相同，在语调判断任务中，被试需要判断句子是陈述语气还是疑问语气，陈述语气按 F 键，疑问语气按 J 键；在声调判断任务中，被试需要判断句子的末尾字是第二声还是第四声，第二声按键盘上的 2 键，第四声按 4。在实验中，首先，被试坐在北师大行为实验室的测试屋里，面对电脑屏幕。主试向被试说明实验要求等指导语，实验开始。每个任务中，都有 10 个练习的 trial，会反馈正确或错误。需要注意的是，对于语气判断任务，这 10 个 trial 中的句子都是陈述语气材料（简称为“s 材料”）或疑问语气材料（简称为“q 材料”），并不涉及换了音高后的材料，因此有正确答案。练习之后有短暂休息，之后是正式的测试阶段，会将 256 个句子放两遍，因此一共有 512 个句子，播放顺序随机。一个任务的正式测验共分为四个 block,每个 block 有 128 个句子，block 之间有休息。每个任务时长半个小时，整个实验一共一个小时左右，整个实验流程都是通过软件 E-prime 来控制的。

此外，在做行为实验的同时，研究者会记录被试的脑电信号，本研究采用 64-channel (Ag/AgCl) NeuroScan 系统的 SynAmps EEG amplifier，带通滤波设置为 0.05－200 Hz，采样率为 1000 Hz.。64 个 channel 位于 Quick-Cap 上，ground 电极位于 FPz 和 Fz 之间，参考电极在鼻尖。垂直眼电(VEOGs)是从左眼上方、下方的一对电极纪录得到的，水平眼电

(HEOGs)是从左眼和右眼的外眼角旁边的位置记录得到。记录中每个电极的电阻保持在10kΩ以下。

### 3.1.5 数据分析

首先计算被试在两个任务中每个条件下的反应正确率，对于语气判断任务中的音高变为疑问语气音高之后的陈述语气材料（简称为“q(s)材料”）与音高变为陈述语气音高之后的疑问语气材料（简称为“s(q)材料”），没有标准答案，为了计算方便，根据被试判断的倾向性，选择被试偏向于参考的一个言语线索（音高或原始语气）作为正确答案。由于数据显示被试整体倾向于将音高变为疑问语气音高之后的陈述语气材料（简称为“q(s)材料”）判断成疑问语气，将音高变为陈述语气音高之后的疑问语气材料（简称为“s(q)材料”）判断成陈述语气，也就是说，被试倾向于依照音高作为判断标准，因此，为了之后进一步的计算方便，研究者将据此作为这两个条件的正确答案的标准。

对于脑电数据，预处理中首先用 NeuroScan 软件[74]对眨眼进行处理，接着对数据进行过滤(low-pass filter of 30 Hz, zero-phase, 24 dB/octave)，并分段（1400ms，刺激前 200ms，刺激开始后 1200ms），以及基线校正。研究者会从 64 个电极中选取出 F3、Fz、F4、C3、Cz、C4、P3、Pz、P4 九个电极，用 Scan 软件算出所有被试在两个任务里的每个刺激类型下每个电极上的平均波幅。按照位置，将九个电极分为左侧、中部、右侧，和额区，中心区，顶区三个区域。每个区域都包含三个电极的数据。



## 3.2 数据结果

### 3.2.1 探究声强线索在语气和声调加工中的作用

因为音高被普遍认为在语气识别中起到了主要作用，本文研究者在评估被试对于不自然刺激的判断的正确率的时候，按照他们的结果是否跟音高线索一致进行判断。分析结果显示，被试对八种刺激类型的语气识别正确率（包括四种不自然刺激）都显著高于随机水平，即 50% (all $ps < 0.005$)，见表 2。考虑到在四种不自然刺激中，声强线索指示了与音高线索相反的语气，基于音高线索进行分析得到的显著高于随机水平的识别率证实了音高（而不是声强）在语气识别中的主要作用。

表 2 语气任务的平均正确率（* $p < 0.00625$, Bonferroni corrected $p < 0.05$）。

| 刺激类型 | 平均正确率 | t 值 | 显著性 |
|---|---|---|---|
| q2 | 0.782 | 5.608 | <0.001* |

| | | | |
|---|---|---|---|
| q4 | 0.954 | 31.382 | <0.001* |
| s2 | 0.924 | 23.208 | <0.001* |
| s4 | 0.977 | 62.394 | <0.001* |
| s2(q) | 0.847 | 14.301 | <0.001* |
| s4(q) | 0.902 | 14.431 | <0.001* |
| q2(s) | 0.657 | 3.394 | 0.005* |
| q4(s) | 0.860 | 9.097 | <0.001* |

以任务（语气任务，声调任务）、音高（疑问音高，陈述音高）、声强（疑问声强、陈述声强）和声调（二声，四声）为被试内变量的重复测量方差分析发现，声强的主效应并不显著（$F(1,12)=2.446, p=0.144, \eta_p^2=0.169$），但是有一个显著的三项交互作用，任务*声强*音高（$F(1,12)=87.161, p<0.001, \eta_p^2=0.879$），见图 4。具体表现为，对于语气任务，声强*音高的交互作用是显著的($F(1,12)=122.884, p<0.001, \eta_p^2=0.911$)，对于疑问音高的刺激类型，疑问声强的刺激比陈述声强的刺激有更高的识别正确率($p<0.001$)；对于陈述音高的刺激类型，陈述声强的刺激比疑问声强的刺激有更高的正确率($p<0.001$)。这一结果表明当声强和音高一致时，比不一致情况更容易识别语气，而这种交互作用只在语气任务里存在，在声调任务里并没有发现。除了与声强相关的效应以外，方差分析还发现了显著的任务主效应（$F(1,12)=57.631, p<0.001, \eta_p^2=0.828$）、音高主效应（$F(1,12)=12.462, p=0.004, \eta_p^2=0.509$）、声调主效应（$F(1,12)=38.594, p<0.001, \eta_p^2=0.763$），以及任务和声调的交互作用（$F(1,12)=51.453, p<0.001, \eta_p^2=0.811$），由于这些效应与声强无关，所以会在接下来探究语气和声调在两个任务中的效应的分析中再讨论。

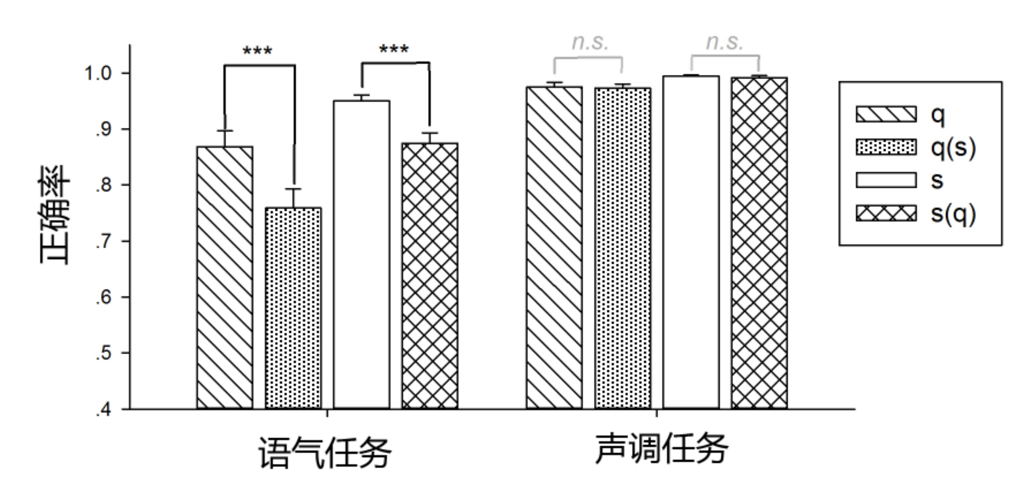



图 4 声强*音高的两项交互作用只在语气识别任务中显著（*n.s.*表示 $p>0.05$，*表示 $p<.05$，**表示 $p<.01$，***表示 $p<.001$；误差线表示标准误）。

以上的行为结果说明音高是在语气识别中最主要也是最重要的线索，不过声强也通过与音高的交互作用对语气加工有所影响。但是在声调识别中，并未发现声强有什么作用。

既然已经考察了声强的重要性，考虑到它对语气和声调识别的非核心的作用，在之后的分析中研究者将移除不自然的刺激，分析被试对于自然刺激的行为和脑电反应以更好地探究语气和声调线索是怎么各自参与到自然的语气识别和声调识别动态过程中的。

### 3.2.2 探究语气和声调在两个任务中的效应

对于行为数据，以任务（语气识别任务，声调识别任务）、语气（疑问语气，陈述语气）和声调（二声，四声）为被试内变量的重复测量方差分析发现了显著的任务主效应 ($F(1,12)=28.215$, $p<0.001$, $\eta_p^2=0.702$)，声调任务比语气任务的正确率高。语气的主效应也显著($F(1,12)=10.166$, $p=0.008$, $\eta_p^2=0.459$)，在两个任务中疑问语气都比陈述语气有更低的正确率（语气任务 $p=0.008$，声调任务 $p=0.043$；图 5）。同时还发现了显著的任务*声调交互作用($F(1,12)=20.911$, $p=0.001$, $\eta_p^2=0.635$)，见图 6，在语气任务中二声结尾的刺激的语气识别率比四声刺激的识别率低($p<0.001$)，但在声调任务中二声和四声之间没有显著的差异($p=0.165$)。

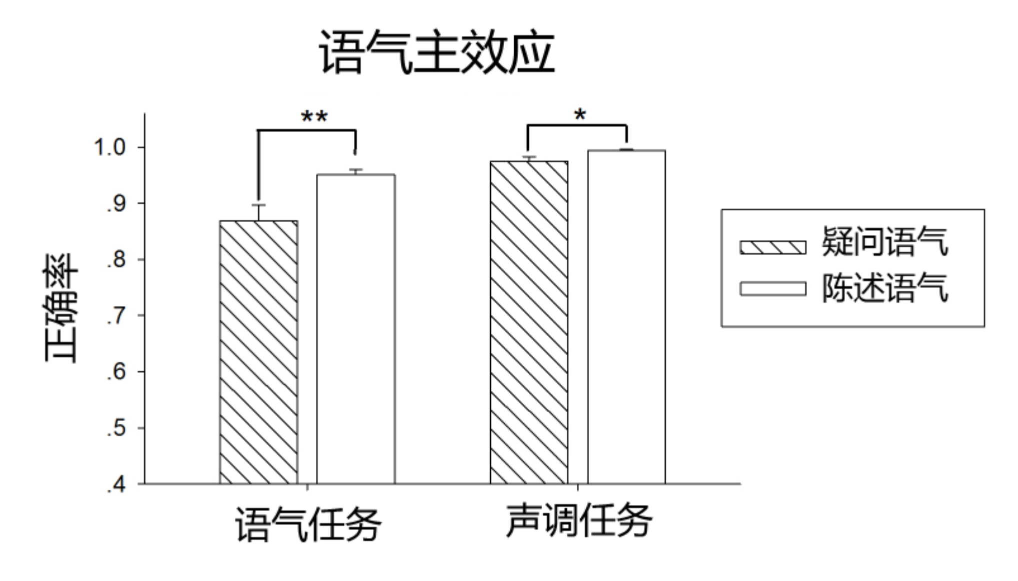



图 5 语气的主效应（*n.s.*表示 $p>0.05$，*表示 $p < .05$，**表示 $p < .01$，***表示 $p < .001$；误差线表示标准误）。

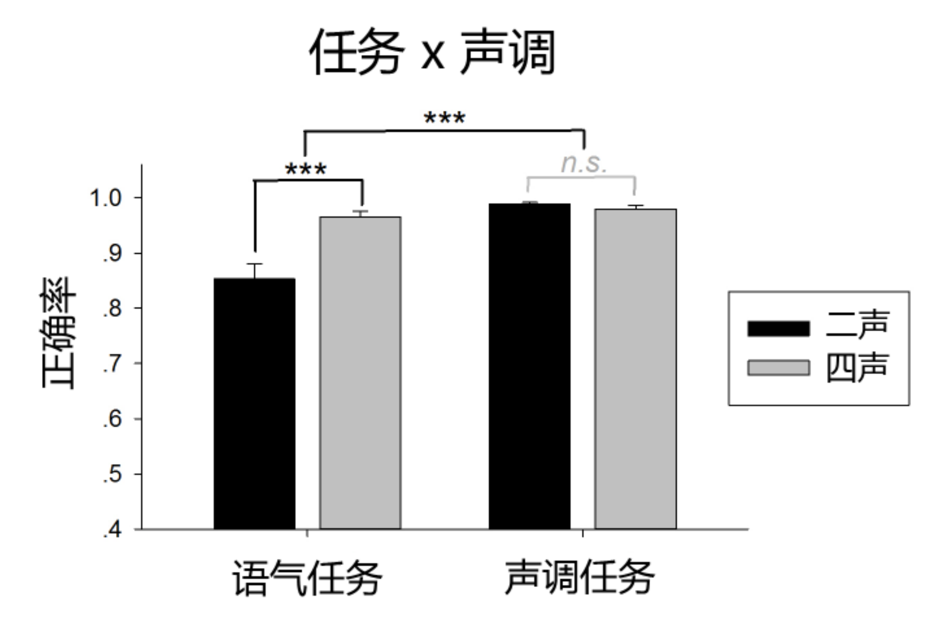



图 6 任务和声调的交互作用（*n.s.*表示 $p>0.05$，*表示 $p<.05$，**表示 $p<.01$，***表示 $p<.001$；误差线表示标准误）。

为了探索末尾字听觉特征与两个任务中的识别率之间的关系，研究者做了一个线性回归分析，包括了 F0 level 和 F0 interval 作为自变量。结果发现对于语气识别，包括了 F0 interval、移除了 F0 level 的回归模型是显著的($p<0.001$)，与之相反，对于声调识别，显著的回归模型包含了 F0 level、没有 F0 interval ($p<0.001$)。这意味着虽然从听觉分析看来 F0 level 和 F0 interval 都可以区分语气和声调，但当人们在听这些包含着语气和声调的自然刺激时，对语气的识别可能主要被 F0 interval 影响，对声调的识别更多与 F0 level 有关。

对于脑电数据，研究者分析了被试从听到关键词开始到 800ms 时这段时间内每 50ms 时间窗里九个电极上波幅的平均值，并对被试在听自然刺激时的脑电信号的平均波幅做了重复测量方差分析，被试内变量包括语气（疑问，陈述）、声调（二声，四声）、区域（额区，中心区，顶区）和半球（左侧，中部，右侧）。50ms 的连续时间窗口分析结果见表 3。在方差分析中得到相同效应的连续 50ms 时间窗会合并在一起，重新进行上述方差分析（只有超过两个连续 50ms 时间窗都发现了的效应才会被讨论）。以下结果是基于合并后的时间窗报告的。

在语气任务中，对 150-250ms 时间窗里的九个电极的平均波幅的方差分析发现了一个显著的语气*区域的交互效应（$F(2,24)=27.958$, $p<0.001$, $\eta_p^2=0.700$），在顶区，疑问语气会比陈述语气诱发一个更大的负成分（$p=0.017$）。其他的效应都不显著。在 250-400ms 也发现了显著的语气*区域的交互效应（$F(2,24)=12.886$, $p<0.001$, $\eta_p^2=0.518$），在顶区，疑问语气比起陈述语气有更大的负成分（$p=0.012$）。同样在这个时间窗，有个显著的声调*区域的效应（$F(2,24)=3.607$, $p=0.043$, $\eta_p^2=0.231$），在顶区，二声比四声诱发一个更大的负成分（$p=0.038$）。在 450-650ms 时间窗，有显著的语气*声调（$F(1,12)=6.733$, $p=0.023$,

$\eta_p^2$=0.359）交互作用，对于四声结尾的刺激，陈述语气比疑问语气有更小的正成分（*p*=0.001）；对于二声结尾的刺激，没有这样两个语气之间平均波幅的显著差异（*p*=0.868）。在 650-750ms 时间窗，有显著的语气*区域（F(2,24)=3.987, *p*=0.032, $\eta_p^2$=0.249）和声调*区域（F(2,24)=7.869, *p*=0.002, $\eta_p^2$=0.396）主效应。在顶区，疑问语气比陈述语气有更大的正成分（*p*=0.031）。在中央区（*p*=0.004）和顶区（*p*=0.002），四声比二声诱发出更大的正成分。图 7 有以上的显著效应，因为大多数效应都出现在顶区，因此在这个图（以及图 8）中呈现的脑电波形图和方差分析显著效应都是顶区（P3, Pz, P4 电极的平均值）的平均波幅。



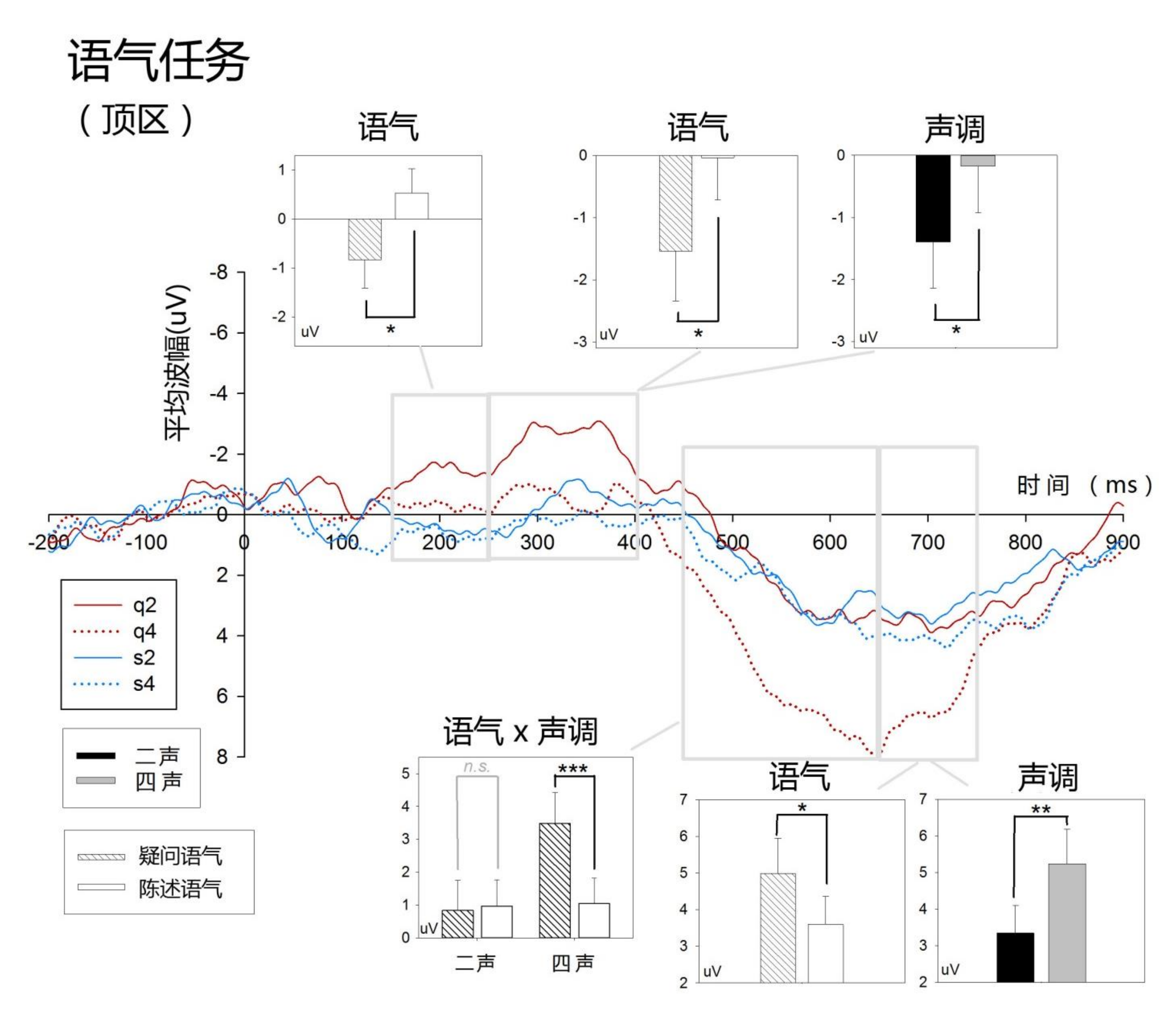



图 7 语气任务中的显著效应（*n.s.*表示 $p$>0.05，*表示 $p < .05$，**表示 $p < .01$，***表示 $p < .001$；误差线表示标准误）。

在声调任务中，100-250ms 有一个显著的语气*声调*区域之间的三项交互作用 (F(2,24)=8.791, *p*=0.001, $\eta_p^2$=0.423)，对于疑问句来说，在顶区，四声有一个比二声更大的

负成分($p$=0.034)，对于陈述句，没有发现声调之间的显著差异($p$>0.05)。其他的效应都不显著。结果见图 8。

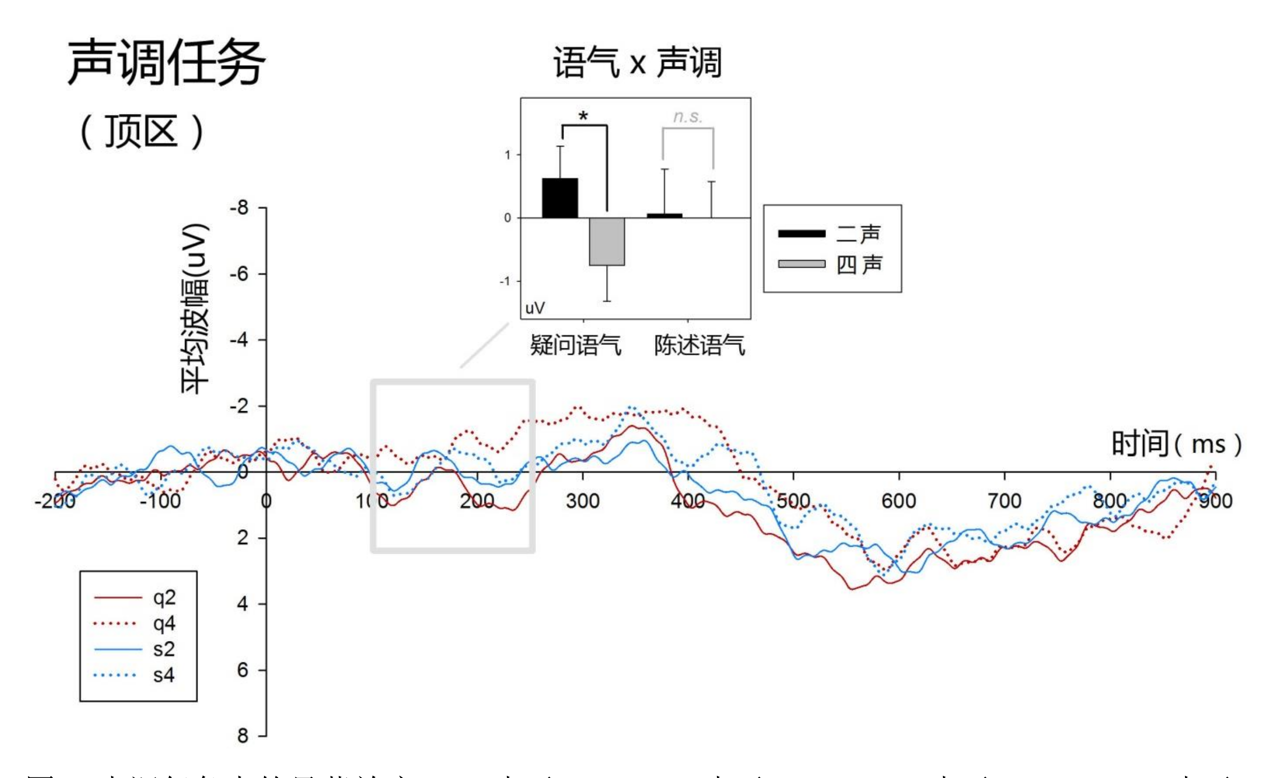



图 8 声调任务中的显著效应（*n.s.*表示 $p$>0.05，*表示 $p < .05$，**表示 $p < .01$，***表示 $p < .001$；误差线表示标准误）。

表 3 所有 50ms 连续时间窗口（0-800ms）方差分析的显著效应（该表呈现了包含语气和声调变量的显著效应；当交互效应包含了某变量，该变量的主效应不再单独列出）。

| 语气识别任务 | | 声调识别任务 | |
|---|---|---|---|
| 时间窗(ms) | 显著效应 | 时间窗(ms) | 显著效应 |
| 0-50 | × | 0-50 | × |
| 50-100 | 语气*声调*区域*半球 | 50-100 | × |
| 100-150 | × | 100-150 | 语气*声调*区域 |
| 150-200 | 语气*区域 | 150-200 | 语气*声调*区域 |
| 200-250 | 语气*区域 | 200-250 | 语气*声调*区域 |
| 250-300 | 语气*区域 | 250-300 | × |
| 300-350 | 语气*区域<br>声调*区域 | 300-350 | × |
| 350-400 | 语气*区域*半球 | | × |
| 400-450 | × | 400-450 | 语气*声调*区域 |
| 450-500 | 语气*声调 | 450-500 | × |

| | | | |
|---|---|---|---|
| 500-550 | 语气*声调 | 500-550 | 声调*区域 |
| 550-600 | 语气*声调 | 550-600 | × |
| 600-650 | 语气*声调 | 600-650 | × |
| 650-700 | 语气*区域<br>声调*区域 | 650-700 | × |
| 700-750 | 声调*区域<br>语气*声调*区域 | 700-750 | × |
| 750-800 | × | 750-800 | 语气*区域*半球 |

## 3.3 讨论

实验一探究了汉语中语气和声调加工的时序层级，研究者记录并分析了被试在做语气识别和声调识别任务时，对在相同音节中含有语气和声调信息的汉语句子的脑电反应。在语气任务中，从脑电平均波幅的结果来看，被试首先对语气有反应(100-250ms 时间窗出现语气主效应)，跟声调有关的效应在 250-400ms 之后才出现。在声调任务中，最先被探察到的是 100-250ms 时间窗出现语气*声调的交互作用，被试对语气和声调通过一个早期的交互作用有所反应：在 100-250ms 时间窗，疑问语气里四声比二声具有更大的负成分，而陈述语气里二四声的平均波幅没有显著差异。可以看到大脑对二声和四声有所区别的脑电反应是依靠于语气特征的，当语气不同时，大脑对声调的反应也不同。如果最初的任务相关假设成立，在声调任务中应该首先看到声调自身的效应，然而实际上语气在早期的时间窗口也有效应，说明语气也在早期就已经参与进了声调任务的加工中，这个结果揭示了被试对语气的敏感性，相对更倾向于第二个假设，语气有更高的时序层级，无论在语气任务还是声调任务中都有早期效应。

与声调任务显著的语气和声调之间的交互作用相类似，在语气任务中，在 450-650ms 时间窗也存在语气和声调显著的交互作用，四声结尾的疑问刺激有比陈述刺激更大的正成分，根据之前的研究结果看来，这可能是由于上升的疑问语气和下降的声调之间的冲突[5, 6, 8-11]。在语气任务接下来的 650-750ms 时间窗中语气和声调的主效应（疑问语气和四声有更大的正成分）很可能也是由于这样的疑问语气和四声之间的冲突导致的。在这一时间窗中发现的正成分可以被视为 P600 成分，它被视为是再加工过程的一个反映[75, 76]，而且也在前人关于声调和语气之间交互的研究中被发现[11]。P600 反映了一个被试为了解决他们探察到的一些冲突和没预料到的事件而重新分析和检查句子的过程[75]。在当前研究中，因为四声和疑问之间的强烈冲突，在做一个识别决策之前，被试可能需要再加工一下有疑问语气的或者是有四声的末尾音节，好解决可能的冲突，避免判断错误。语气任务和声调任务中出现的这两个交互作用并没有在行为结果中出现，可能是由于识别任务本身的天花板效应。在声调任务中，对疑问二声、疑问四声、陈述二声、陈述四声的

刺激的平均声调识别率为 0.98, 0.97, 0.99 和 0.99（标准差分别是 0.02、0.05、0.01 和 0.01），在语气任务中，虽然疑问二声刺激的识别率要低于其他几个条件（识别率为 0.78，标准差为 0.18），疑问四声、陈述二声、陈述四声刺激的识别率为 0.95、0.92 和 0.98（标准差分别为 0.05、0.07 和 0.03）。这两个任务可能对于这些汉语母语的被试来说太简单了，所以有些效应显示在脑电结果里，但在行为数据上没有表现。脑电的结果可能是由自上而下的加工和自下而上的加工一起导致的。关于跟刺激的听觉属性有关的自下而上的加工，F0 level 被发现在对声调识别率做的回归模型中显著，而 F0 interval 在语气的回归模型中显著，但语气和声调之间的交互作用无论是 F0 level 或是 interval 都没法解释。正如同在图 3 中显示的一样，四声的 F0 level 无论在疑问还是陈述语气中都比二声更高，F0 interval 也是。因此，这个交互作用无法由刺激的自下而上加工所解释。考虑自上而下的调节，在声调识别任务中疑问四声刺激更大的负波可以被疑问语气和四声之间的冲突所解释，这种冲突通过变平四声的下降的斜率，很大地改变了四声的 F0 轮廓(见图 1；也可参考 Liu 等人 2016 年的研究[10])，使得对声调的识别变得更困难了。

除了语气和声调的交互作用，在语气任务中还发现了显著的主效应，在 100-250ms 和 250-400ms，疑问都比陈述语气有更大的负成分，这可以从疑问比陈述更难识别的行为结果得以解释。如 Yuan 所说[5]，对于陈述语气的识别更简单，可能因为陈述语气是一种默认语气，而疑问语气是标记语气，需要一些特定的特征来识别它。在 250-400ms 二声有比四声更大的负成分可能是因为在识别语气时末尾是二声时比四声时有更大的难度。

本研究的大部分脑电发现都集中在顶叶区域，这与之前关于语气和声调的效应在中央-顶叶区的发现是一致的[10, 11]。在 Liu 和同事的研究中，四声结尾的疑问和陈述刺激在后部脑区比中央区域更显著[10]。在前人关于言语加工的功能性磁共振研究中发现了一个声调加工的背侧额顶网络[77]，在另一个研究中，当将主动条件（被试关注语气或声调并做出辨别判断）与被动听刺激的条件进行对比时，也发现了顶叶的激活[78]。综上，顶叶可能在语气和声调的加工中扮演了一个重要的角色，前人的磁共振研究[79]发现顶叶皮层在被试观看社会交往相关的视频片段时会有所激活，顶叶的社会认知功能很可能与本研究发现的顶叶在语气和声调任务上的参与相关。

据我们所知，当前研究为大脑对同一个言语音节中的语气和声调的加工时序的探索首次提供了证据。对于语气在语气和声调识别加工过程中的早期效应的发现突出了语气以及它所承载的社会意义对人类社会性交流的重要性。

# 第 4 章 实验二：音乐障碍对语气与声调加工的影响

## 4.1 实验方法

### 4.1.1 被试样本

首先，研究者通过在校园论坛发被试招募帖子，除了实验一中的 13 个控制组被试（5 男 8 女）以外，本研究还招募到 11 个失乐症成人被试（3 男 8 女）。所有的被试的母语都是汉语普通话，没有接受过超过两年的正规音乐训练，也没有任何神经或者心理疾病的相关病史。听力检测结果显示，所有被试在 250 至 8000 赫兹（Hz）的最小听力阈值均低于 20 分贝（dB），均具有良好听力。利用艾丁伯格利手问卷（Edinburgh Handedness Inventory）[72]测得所有被试都是右利手。他们都完成了蒙特利尔失乐症成套测验（MBEA）[52]以鉴定是否存在失乐症或者声调障碍，测试包括三个音高相关的分测试、两个时间相关的分测试和一个记忆分测试。纯失乐症被试的 MBEA 得分都低于鉴定标准 71.7%，该鉴定标准根据控制组的得分减去两个标准差所得[47]，而控制组被试都具有正常的音乐能力（高于鉴定标准 71.7%）。此外，这些被试在音乐能力筛查测试时也进行了声调能力的测试（声调测试见[47]），任务包含了声调的识别和区分，两组在声调测试上没有显著差异（$p$=0.773）。同时，所有被试都经过了中国版韦氏成人智力量表（WAIS-RC: The Wechsler Adult Intelligence Scale-Revised by China)[80]的测试，两组在智力上没有组间差异，见表 4。

此外，本实验还招募了 76 个儿童被试，他们都是同一个年级（小学三年级）的学生，来自两所北京的小学。他们的音乐能力经过了 MBEMA（The Montreal Battery of Evaluation of Musical Abilities）[81]测试，这一测试包括了音阶、轮廓、音程、节奏和记忆的分测试。测试成绩前 30%的儿童被选为儿童控制组，而成绩后 30%的儿童音乐能力较差，被视为儿童音高障碍组。总共有 21 个儿童在儿童音高障碍组里（10 男 11 女），23 个儿童在儿童控制组里（16 男 7 女）。对这些儿童也进行了智力测试，使用的是英国心理学家 John Carlyle Raven 在 1938 年编制的瑞文标准推理测验（Raven's Standard Progressive Matrices）[82]，结果发现这两组在智力上没有显著差异（$p > 0.10$），同时两组儿童在性别和年龄上都没有显著差异（$ps > 0.10$），见表 5。本研究通过了北京师范大学的伦理审查委员会的授权，所有被试都交了知情同意的书面证明。

表 4 成人控制组和失乐症的人口统计学数据（这两组在年龄、智商、MBEA、声调测试上的组间差异的 p 值都由 one-way ANOVA 得到，性别组间差异的 p 值由 Chi-Square Test 得到。SD 表示标准差；***表示 $p < .001$）。

| | 成人控制组 (n=13) | 成人失乐症 (n=11) | 组间差异 (*p*) |
|---|---|---|---|
| 平均年龄(SD) | 22.8 (2.2) | 22.8 (2.6) | 0.978 |
| 男/女 | 5/8 | 3/8 | 0.562 |
| 智力水平 (SD) | 128.8 (8.4) | 125.0 (9.2) | 0.297 |
| MBEA (SD) | 88.7 (5.1) | 62.8 (6.4) | <0.001*** |
| 声调测试 (SD) | 97.6 (2.6) | 97.3 (3.4) | 0.773 |

表 5 儿童控制组和失乐症的人口统计学数据（这两组在年龄、智商、MBEA 上的组间差异的 p 值都由 one-way ANOVA 得到，性别组间差异的 p 值由 Chi-Square Test 得到。SD 表示标准差；***表示 $p < .001$）。

| | 儿童控制组 (n=23) | 儿童音高障碍组 (n=21) | 组间差异 (*p*) |
|---|---|---|---|
| 平均年龄(SD) | 9.2 (0.4) | 9.0 (0.3) | 0.138 |
| 男/女 | 16/7 | 10/11 | 0.139 |
| 智力水平 (SD) | 44.6 (4.8) | 41.8 (8.0) | 0.155 |
| MBEMA (SD) | 95.6 (2.1) | 78.8 (4.6) | <0.001*** |

#### 4.1.2 刺激材料

为了探究成人失乐症的语气和声调识别能力，实验二使用了与实验一相同的刺激材料。而对于儿童被试，为了减少总的测试时长、保证儿童在实验中的关注和施测质量，研究者在实验一材料中的 32 个刺激中随机选出了一半的刺激作为语气任务中的实验材料，剩下一半作为声调任务中的材料。对刺激的分析发现这两部分材料在声强（t(246)=-0.446, *p*=0.656）、F0 level（t(254)=-0.352, *p*=0.725）、F0 interval（t(254)=-0.527, *p*=0.599）、音长（t(249)=-1.110, *p*=0.268）和词频（t(234)=0.03, *p*=0.976）上都没有显著差异。

#### 4.1.3 刺激呈现

与实验一相同。

### 4.1.4 实验流程

被试会完成一个语气识别任务和一个声调识别任务，做两个任务的顺序随机。对于成人失乐症，与实验一相同，每个句子都播放了两遍，所有句子的播放顺序随机。对于儿童，为了减少测试时间，每个句子只播放一遍。其中一个儿童音高障碍被试由于时间问题没有参加语气任务，其他所有被试都顺利完成了所有任务。

### 4.1.5 数据分析

为了探究每组被试在这八个实验材料条件上的判断正确率是否显著高于随机水平，研究者做了正确率与随机水平（50%）的 t 检验。与实验一相同，对于不自然刺激，由于大部分被试都按照音高进行判断，所以在数据处理时把按照音高判断的数据视为正确。为了研究有音高障碍的被试和控制组被试在完成这两个任务时是否有正确率上的组间差异，并探究不同因素对他们判断的影响，实验中会进行重复测量方差分析，将被试间变量设为组别，被试内变量为任务（语气任务，声调任务）、声强（陈述声强，疑问声强）、音高（陈述音高，疑问音高）和声调（二声，四声）。

## 4.2 数据结果

### 4.2.1 成人控制组与成人失乐症组的结果

研究者对于这两组在语气任务和声调任务上的识别正确率做了与随机水平（50%）对比的单样本 t 检验，在语气任务中，成人控制组对于八种刺激的表现都显著高于随机水平，但成人失乐症被试在判断以二声结尾的疑问语气刺激（即 q2 与 q2(s)条件）时，他们的表现并没有显著区别于随机水平（表 6）。首先，对于自然材料 q2，失乐症没法按照音高或是别的语气线索正确识别出疑问语气，说明二声这个声调显著干扰到了失乐症对疑问语气的判断；其次，对于不自然刺激 q2(s)，失乐症既没有显著地按照音高进行语气识别，也没有利用诸如声强的线索有倾向地按照原来的语气进行判断，这说明二声也影响了对这种不自然刺激的语气识别（无论被试是按照什么线索进行语气识别的）。将失乐症对 q2 和 q2(s)按照音高的识别率进行配对样本 t 检验比较，发现 q2(s)的识别率显著低于 q2 条件（$p$=0.004），说明虽然对这两种刺激失乐症都没法识别出明显的语气来，在音高和声强不一致的情况下的时候失乐症会更不倾向于按照音高进行判断，也就是说，除了声调，声强也影响了语气的识别。在声调任务中，这两组在八种刺激上的判断表现都显著高于随机水平（表 7）。

表 6 成人控制组与成人失乐症在语气任务中的正确率（*代表 $p < 0.00625$，即 Bonferroni corrected $p < 0.05$）。

| | 成人控制组 | | | | 成人失乐症 | | | |
|---|---|---|---|---|---|---|---|---|
| | N | Mean | t | Sig. | N | Mean | t | Sig. |
| q2 | 13 | 0.78 | 5.61 | <0.001* | 11 | 0.66 | 2.36 | **0.040** |
| q4 | 13 | 0.95 | 31.38 | <0.001* | 11 | 0.88 | 6.87 | <0.001* |
| s2 | 13 | 0.92 | 23.21 | <0.001* | 11 | 0.89 | 11.68 | <0.001* |
| s4 | 13 | 0.98 | 62.39 | <0.001* | 11 | 0.89 | 9.00 | <0.001* |
| s2(q) | 13 | 0.85 | 14.30 | <0.001* | 11 | 0.79 | 6.32 | <0.001* |
| s4(q) | 13 | 0.90 | 14.43 | <0.001* | 11 | 0.83 | 7.46 | <0.001* |
| q2(s) | 13 | 0.66 | 3.39 | 0.005* | 11 | 0.54 | 0.65 | **0.528** |
| q4(s) | 13 | 0.86 | 9.10 | <0.001* | 11 | 0.76 | 4.05 | 0.002* |

表 7 成人控制组与成人失乐症在声调任务中的正确率（*代表 $p < 0.00625$，即 Bonferroni corrected $p < 0.05$）。

| | 成人控制组 | | | | 成人失乐症 | | | |
|---|---|---|---|---|---|---|---|---|
| | N | Mean | t | Sig. | N | Mean | t | Sig. |
| q2 | 13 | 0.98 | 96.80 | <0.001* | 11 | 0.99 | 82.21 | <0.001* |
| q4 | 13 | 0.97 | 33.62 | <0.001* | 11 | 0.92 | 10.16 | <0.001* |
| s2 | 13 | 0.99 | 175.54 | <0.001* | 11 | 0.98 | 34.17 | <0.001* |
| s4 | 13 | 0.99 | 175.54 | <0.001* | 11 | 0.96 | 16.55 | <0.001* |
| s2(q) | 13 | 0.99 | 225.48 | <0.001* | 11 | 0.97 | 27.33 | <0.001* |
| s4(q) | 13 | 0.99 | 75.62 | <0.001* | 11 | 0.94 | 11.32 | <0.001* |
| q2(s) | 13 | 0.98 | 91.25 | <0.001* | 11 | 0.97 | 35.08 | <0.001* |
| q4(s) | 13 | 0.96 | 36.32 | <0.001* | 11 | 0.90 | 7.00 | <0.001* |

研究者对两个任务的正确率做了重复测量方差分析，被试间变量为组别（成人失乐症，成人控制组），被试内变量为任务（语气任务，声调任务）、声强（陈述声强，疑问声强）、音高（陈述音高，疑问音高）和声调（二声，四声）。结果发现，语气任务比声调任务显著更难（$F(1,22)=102.839, p<0.001, \eta_p^2=0.824$），在两个任务中整体而言疑问音高比陈述音高更难判断（$F(1,22)=18.143, p<0.001, \eta_p^2=0.452$）。显著的任务*声强*音高的交互效应（$F(1,22)=65.684, p<0.001, \eta_p^2=0.749$）显示，声强和音高的不一致只会对语气判断造成影响。当声强和音高线索不一致时，在语气任务，被试的判断显著差于当声强和音高线索一致的情况（$ps<0.001$；图 9）。相反，在声调任务，一致条件和不一致条件没有显著的差异（$p>0.05$；图 10）。任务*声调交互作用显著($F(1,22)=91.289, p<0.001, \eta_p^2=0.806$)，任务*音

高*声调的交互作用也显著（F(1,22)=16.564, $p$=0.001, $\eta_p^2$=0.430），如图 11 所示，在语气任务中，二声比四声的刺激更难判断出疑问语气($p$<0.001)，而在声调任务中，对于疑问音高的刺激，四声比二声更难被判断出声调($p$=0.026)。这表示四声和疑问音高的组合会让疑问语气变得更好识别，但同时却使四声的声调更难被判断出来了。

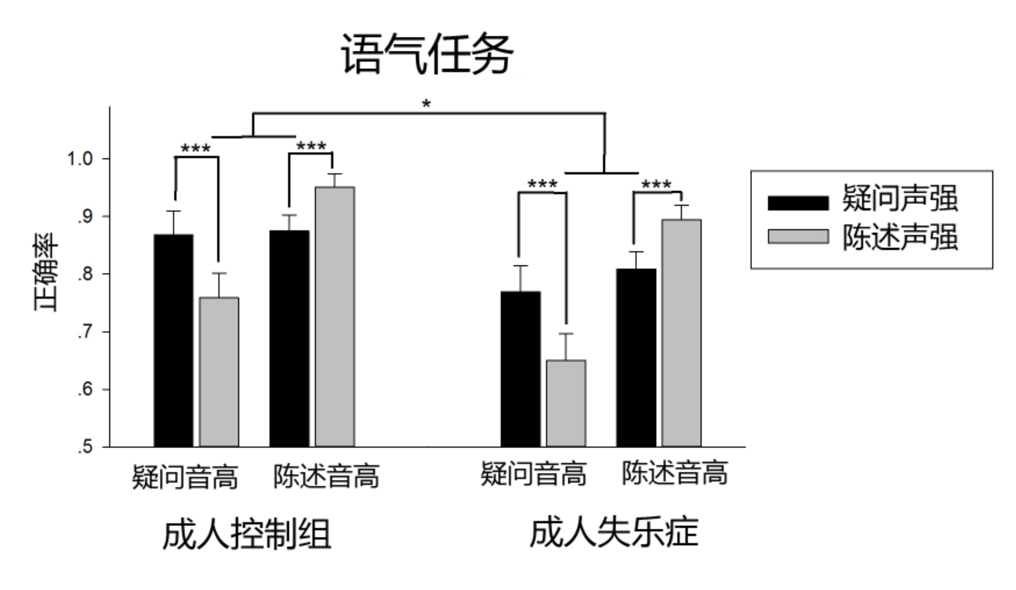
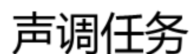



图 9 语气任务的组间差异和音高*声强交互作用（在语气任务中，成人失乐症表现显著差于成人控制组。当声强和音高线索不一致时，被试的判断显著差于当声强和音高线索一致的情况。*n.s.*表示 $p$>0.05，*表示 $p$ < .05，**表示 $p$ < .01，***表示 $p$ < .001；误差线表示标准误）。



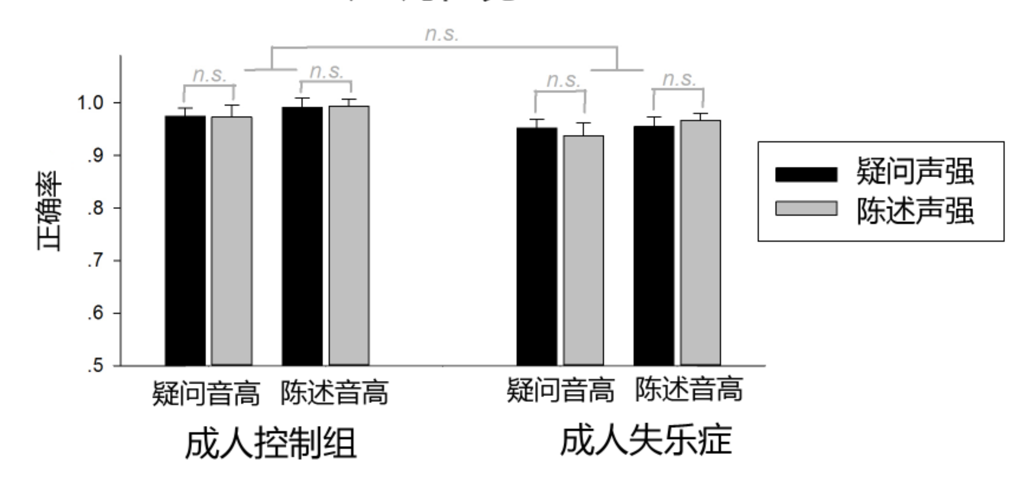



图 10 声调任务（声调任务没有组间差异；音高和声强一致条件和不一致条件之间没有显著的正确率差异。*n.s.*表示 $p$>0.05，*表示 $p$ < .05，**表示 $p$ < .01，***表示 $p$ < .001；误差线表示标准误）。

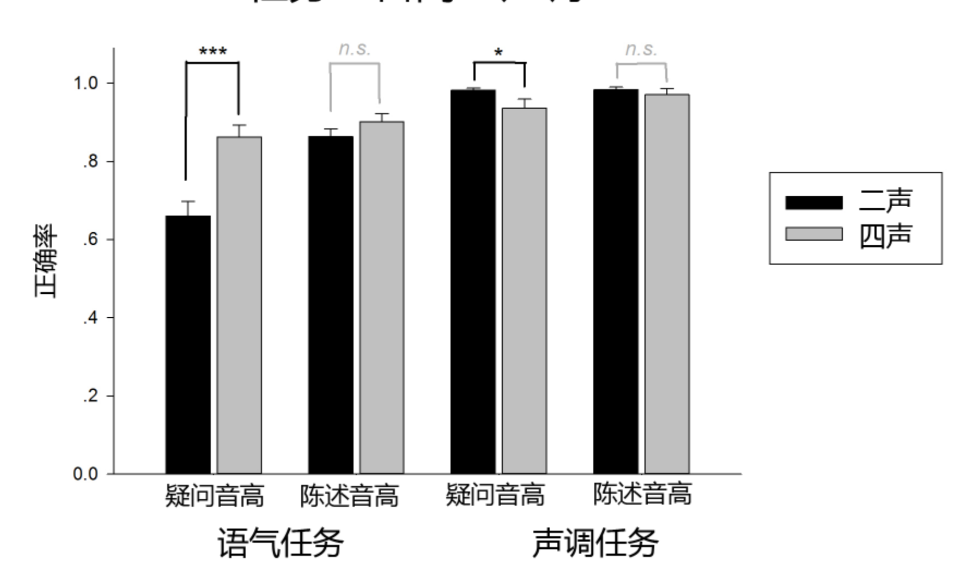



图 11 任务*音高*声调的交互作用（四声使疑问语气更易识别，而疑问语气让四声更难识别。*n.s.*表示 $p>0.05$，*表示 $p<.05$，**表示 $p<.01$，***表示 $p<.001$；误差线表示标准误）。



为了探究在语气和声调任务内部两组是否有组间差异，研究者也对这两个任务分别做了方差分析，被试间变量是这两组的组别，被试内变量是声强、音高和声调。在语气任务中，发现显著的组别主效应($F(1,22)=4.363, p=0.049, \eta_p^2=0.166$；图 9)，但在声调任务中，失乐症并没有显著差于控制组($F(1,22)=1.549, p=0.226$；图 10)。这表示失乐症在音高感知上的缺陷对语气识别有负面影响，但并没有阻碍他们的声调识别。

#### 4.2.2 儿童控制组与儿童音高障碍组的结果

在语气任务中，单样本 t 检验的结果显示两组儿童都能够成功识别大部分的语气刺激（见表 8）。儿童音高障碍组在判断原本是陈述语气、音高换为疑问语气音高的四声结尾刺激（即 q4(s)）时，正确率没有显著高于随机水平。而在声调任务中，无论是儿童控制组还是儿童音高障碍组都不能成功判断出含有疑问音高和四声的两个刺激条件（q4 和 q4(s)）的声调（见表 9）。声调任务的这个 t 检验结果说明，整体而言，疑问的音高都严重干扰了儿童对四声的声调识别，不论声强等其他线索是什么，也不论儿童是否存在音乐障碍。而在语气任务中，儿童音高障碍组在 q4(s)上无法识别语气，这说明，四声和陈述声强的线索阻碍了有音高加工障碍的儿童按照疑问音高进行语气识别，但与此同时，他们对于自然条件 q4 的语气判断是显著高于随机水平的，这进一步说明，四声本身不会使儿童音高障碍组判断不出疑问语气，而是不一致的声强等线索加大了他们对这种四声

加疑问音高的刺激进行语气识别的困难，造成他们最终无法判断出语气的结果。这个结果说明了儿童音高障碍组在语气加工时对于声强线索的倚重，这可能与他们在音高上的障碍从而导致的在其他线索上的补偿策略有关。

表 8 儿童控制组与儿童音高障碍组在语气任务中的正确率（*代表 $p < 0.00625$，即 Bonferroni corrected $p$ <0.05）。

| | 儿童控制组 | | | | 儿童音高障碍组 | | | |
|---|---|---|---|---|---|---|---|---|
| | N | Mean | t | Sig. | N | Mean | t | Sig. |
| q2 | 23 | 0.80 | 9.80 | <0.001* | 20 | 0.70 | 4.09 | 0.001* |
| q4 | 23 | 0.88 | 15.94 | <0.001* | 20 | 0.73 | 3.58 | 0.002* |
| s2 | 23 | 0.79 | 9.00 | <0.001* | 20 | 0.66 | 4.28 | <0.001* |
| s4 | 23 | 0.92 | 20.71 | <0.001* | 20 | 0.87 | 15.40 | <0.001* |
| s2(q) | 23 | 0.71 | 5.51 | <0.001* | 20 | 0.64 | 4.18 | 0.001* |
| s4(q) | 23 | 0.83 | 9.99 | <0.001* | 20 | 0.75 | 9.65 | <0.001* |
| q2(s) | 23 | 0.76 | 11.35 | <0.001* | 20 | 0.70 | 4.82 | <0.001* |
| q4(s) | 23 | 0.75 | 5.88 | <0.001* | 20 | 0.59 | 1.54 | **0.141** |

表 9 儿童控制组与儿童音高障碍组在声调任务中的正确率（*代表 $p < 0.00625$，即 Bonferroni corrected $p$ <0.05）。

| | 儿童控制组 | | | | 儿童音高障碍组 | | | |
|---|---|---|---|---|---|---|---|---|
| | N | Mean | t | Sig. | N | Mean | t | Sig. |
| q2 | 23 | 0.93 | 29.39 | <0.001* | 21 | 0.88 | 16.74 | <0.001* |
| q4 | 23 | 0.67 | 2.86 | **0.009** | 21 | 0.57 | 1.18 | **0.251** |
| s2 | 23 | 0.94 | 20.00 | <0.001* | 21 | 0.88 | 14.20 | <0.001* |
| s4 | 23 | 0.90 | 22.99 | <0.001* | 21 | 0.87 | 14.92 | <0.001* |
| s2(q) | 23 | 0.92 | 28.88 | <0.001* | 21 | 0.84 | 11.96 | <0.001* |
| s4(q) | 23 | 0.91 | 15.43 | <0.001* | 21 | 0.84 | 9.36 | <0.001* |
| q2(s) | 23 | 0.94 | 25.26 | <0.001* | 21 | 0.85 | 11.58 | <0.001* |
| q4(s) | 23 | 0.67 | 2.63 | **0.015** | 21 | 0.51 | 0.23 | **0.817** |

重复测量方差分析（被试间变量是组别，被试内变量是任务、音高、声强和声调）发现了对儿童来说，声调任务也比语气任务更简单（$F(1,41)=19.156, p<0.001, \eta_p^2=0.318$）。有显著的任务*音高的交互作用（$F(1,41)=7.684, p=0.008, \eta_p^2=0.158$），在声调任务里，疑问音高比陈述音高更难判断声调（$p<0.001$），在语气任务里，两个音高在语气判断上没有显著正确率差异（$p=0.264$）。任务*音高*声调的交互作用显著（$F(1,41)= 7.440, p=0.009, \eta_p^2=0.154$），

在语气任务里，对于陈述音高，二声比四声更难判断语气（$p$<0.001），对于疑问音高，二声四声没有显著差异（$p$=0.949）；在声调任务里，对于疑问音高，四声比二声更难判断声调（$p$<0.001），对于陈述音高，二四声没有显著差异（$p$=0.296）。

此外，四项交互作用任务*声强*音高*组别显著（$F(1,41)=4.309$, $p=0.044$, $\eta_p^2=0.095$），如图 12 所示。在语气任务，对于自然刺激和不自然刺激，儿童音高障碍组的表现都显著差于儿童控制组($ps < 0.05$)，而在声调任务，只在不自然刺激中发现了音高障碍的儿童比控制组儿童更差的结果($ps < 0.05$)。两组儿童在对自然刺激进行声调判断上没有显著组间差异，而在语气判断时有，这说明音高加工障碍对于儿童的语气判断的影响比对声调判断的影响大。

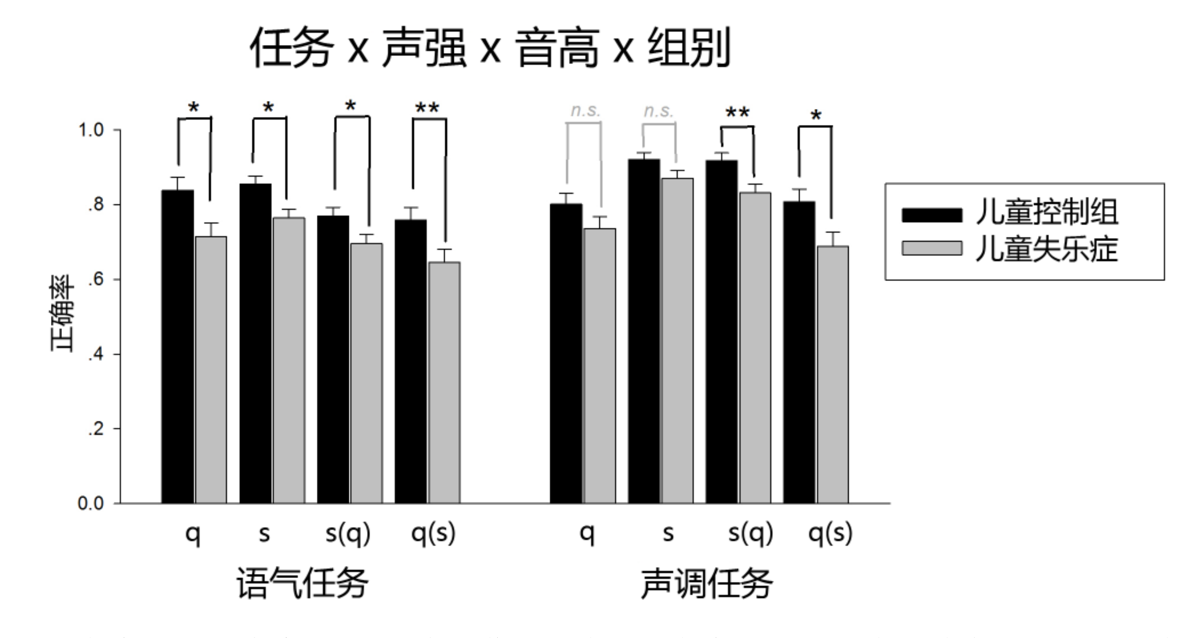



图 12 任务*声强*音高*组别的交互作用（在声调任务的自然刺激正确率上，两组没有显著组间差异。*n.s.*表示 $p$>0.05，*表示 $p < .05$，**表示 $p < .01$，***表示 $p < .001$；误差线表示标准误）。



#### 4.2.3 成人控制组与儿童控制组的比较

t 检验的结果表明对于儿童和成人的控制组，八种刺激类型的语气识别率都显著高于随机水平（即 50%），见表 10。在声调识别上，成人控制组显著高于随机水平，但儿童控制组却不能够成功识别出疑问音高的四声刺激（即 q4 和 q4(s)条件，见表 11）。儿童控制组能够成功判断所有语气刺激却不能判断出一些条件下的声调的发现揭示了儿童已经具备依靠音高线索判断语气的能力，但判断声调的能力还没有完全发展成熟。

表 10 成人控制组与儿童控制组在语气任务中的正确率（*代表 $p < 0.00625$, 即 Bonferroni

corrected $p$ <0.05）。

| | 成人控制组 | | | | 儿童控制组 | | | |
|---|---|---|---|---|---|---|---|---|
| | N | Mean | t | Sig. | N | Mean | t | Sig. |
| q2 | 13 | 0.78 | 5.61 | <0.001* | 23 | 0.80 | 9.80 | <0.001* |
| q4 | 13 | 0.95 | 31.38 | <0.001* | 23 | 0.88 | 15.94 | <0.001* |
| s2 | 13 | 0.92 | 23.21 | <0.001* | 23 | 0.79 | 9.00 | <0.001* |
| s4 | 13 | 0.98 | 62.39 | <0.001* | 23 | 0.92 | 20.71 | <0.001* |
| s2(q) | 13 | 0.85 | 14.30 | <0.001* | 23 | 0.71 | 5.51 | <0.001* |
| s4(q) | 13 | 0.90 | 14.43 | <0.001* | 23 | 0.83 | 9.99 | <0.001* |
| q2(s) | 13 | 0.66 | 3.39 | 0.005* | 23 | 0.76 | 11.35 | <0.001* |
| q4(s) | 13 | 0.86 | 9.10 | <0.001* | 23 | 0.75 | 5.88 | <0.001* |

表 11 成人控制组与儿童控制组在声调任务中的正确率（*代表 $p < 0.00625$，即 Bonferroni corrected $p$ <0.05）。

| | 成人控制组 | | | | 儿童控制组 | | | |
|---|---|---|---|---|---|---|---|---|
| | N | Mean | t | Sig. | N | Mean | t | Sig. |
| q2 | 13 | 0.98 | 96.80 | <0.001* | 23 | 0.93 | 29.39 | <0.001* |
| q4 | 13 | 0.97 | 33.62 | <0.001* | 23 | 0.67 | 2.86 | **0.009** |
| s2 | 13 | 0.99 | 175.54 | <0.001* | 23 | 0.94 | 20.00 | <0.001* |
| s4 | 13 | 0.99 | 175.54 | <0.001* | 23 | 0.90 | 22.99 | <0.001* |
| s2(q) | 13 | 0.99 | 225.48 | <0.001* | 23 | 0.92 | 28.88 | <0.001* |
| s4(q) | 13 | 0.99 | 75.62 | <0.001* | 23 | 0.91 | 15.43 | <0.001* |
| q2(s) | 13 | 0.98 | 91.25 | <0.001* | 23 | 0.94 | 25.26 | <0.001* |
| q4(s) | 13 | 0.96 | 36.32 | <0.001* | 23 | 0.67 | 2.63 | **0.015** |

包括了两个任务的重复测量方差分析发现，组别的主效应显著（$F(1,34)=16.448$, $p<0.001$, $\eta_p^2=0.326$），组别*音高*任务的交互作用也显著（$F(1,34)=13.053$, $p=0.001$, $\eta_p^2=0.277$），在语气任务中，对于疑问音高的语气，成人控制组和儿童控制组没有识别率上的显著差异（$p=0.674$），但对于陈述音高的语气，儿童控制组的正确率显著差于成人控制组（$p=0.002$）；在声调任务中，不论是疑问还是陈述音高的刺激，儿童控制组的声调识别都显著差于成人控制组（$ps<0.001$）。在语气任务中成人和儿童在陈述语气的识别上的组间差异会在下一段进行详细分析。

语气任务的重复测量方差分析显示，儿童控制组的语气判断正确率显著低于成人控制组（$F(1,34)=5.251$, $p=0.028$, $\eta_p^2=0.134$）。此外，以往研究[5]发现，陈述语气比疑问语气更容易识别，可能是由于听者在找不到足够的线索证明语气是疑问语气的时候，就会把它识别为陈述语气（陈述语气作为一种默认或无标记的语气类型）。在本研究的语气任务中，对于

成人控制组的方差分析（音高、声强和声调作为被试内变量）发现了音高的主效应($F(1,12)=11.663$, $p=0.005$, $\eta_p^2=0.493$)。然而，在儿童控制组中并没有发现显著的音高主效应($F(1,22)=0.250$, $p=0.622$)，如图 13 所示，这说明虽然儿童能够在音高线索的基础上对语气进行判断，但他们还没有形成陈述语气偏向（statement bias）。不过，陈述语气偏向的缺失并不意味着儿童的语气判断能力没有发育成熟，因为这种偏向可能仅仅是一种语气加工时的策略，而不是一个必须具备的能力。

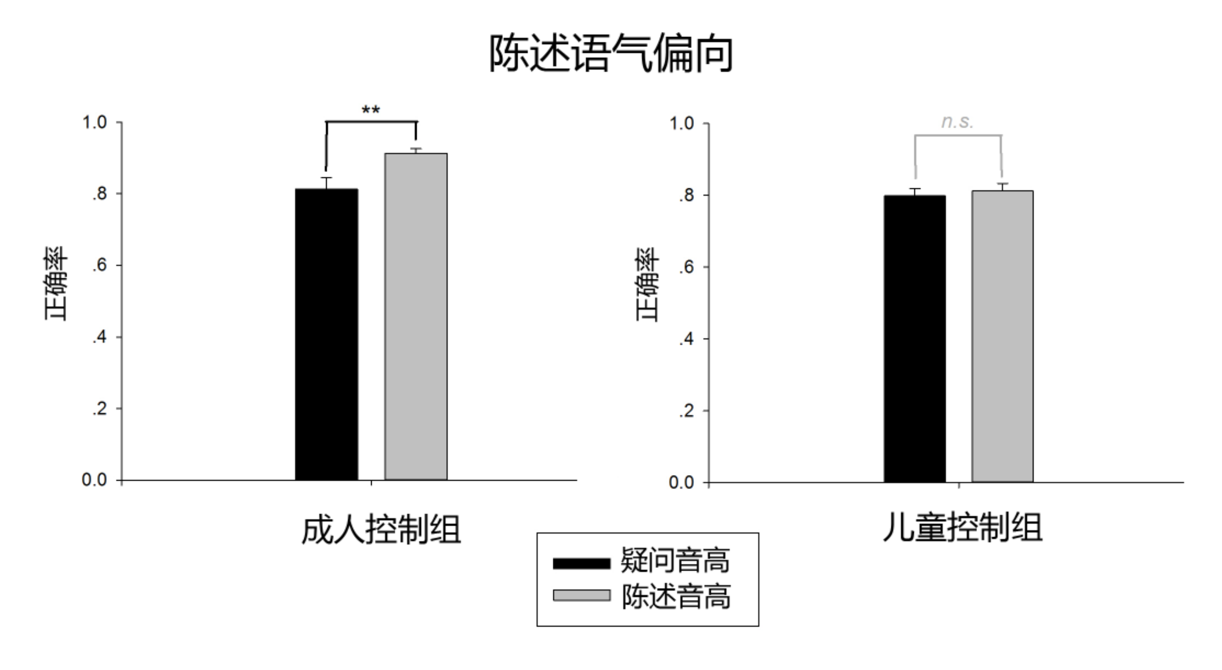



图 13 成人控制组被试的陈述语气偏向（*n.s.*表示 $p>0.05$，*表示 $p<.05$，**表示 $p<.01$，***表示 $p<.001$；误差线表示标准误）。

在声调任务中，以音高、声强和声调作为被试内变量，组别作为被试间变量的方差分析也发现了儿童控制组的声调识别率显著低于成人控制组（$F(1,34)=23.994$, $p<0.001$, $\eta_p^2=0.414$），同时还发现了显著的音高*声调*组别的三项交互作用($F(1,34)=7.499$, $p=0.010$, $\eta_p^2=0.181$；图 14)，只对儿童音高障碍组，当刺激是疑问音高时，四声的正确率显著差于二声($p<0.001$)，在成人控制组中并没有发现这样的音高和声调间的交互作用，这进一步支持了前文的发现：成人已经掌握了的判断疑问音高时的四声声调的能力，在九岁儿童群体（甚至是音乐能力高的儿童）中还没有发展成熟。

图 14 音高*声调*组别三项交互作用（对于儿童控制组，当刺激是疑问音高时，四声的正确率显著差于二声。*n.s.*表示 $p>0.05$，*表示 $p < .05$，**表示 $p < .01$，***表示 $p < .001$；误差线表示标准误）。

## 4.3 讨论

实验二研究了音乐障碍对语气和声调识别的影响，发现在成年人被试群体中，音乐障碍被试在语气任务上表现显著差于正常被试，在声调任务中则与正常被试没有显著差异。而在儿童中，音乐能力差的被试在语气任务上表现出了更低的正确率，在声调任务中，虽然整体正确率也较低，但对于自然刺激的声调识别率与正常被试也是没有显著差异的。结果显示出音乐障碍主要影响了语气识别，而在声调识别上没有太大影响。

成人失乐症在语气上的问题与前人研究的结果相一致[64, 69]，发现音乐障碍会影响语气的识别。t 检验发现成人控制组在判断语气时都按照音高进行判断，成人失乐症也基本上是按照音高进行判断，但失乐症在识别材料是二声时的疑问音高上会出现问题（无论声强是什么）。对于声调的识别，成人无论是控制组还是失乐症都显著高于随机水平，而且不存在组间差异。此外，不论对于成人失乐症还是控制组，在进行语气识别时，当声强和音高线索不一致时，被试的判断都会差于当声强和音高线索一致的情况。而在进行声调识别时，声强和音高一致/不一致的刺激并没有产生正确率上的不同，说明声强的操作对两组都只影响了语气识别、不影响声调识别。

而在儿童群体中，从显著的四项交互作用任务*声强*音高*组别可以看到，在语气识别上，对于自然刺激，儿童音高障碍组差于控制组；对于不自然刺激，儿童音高障碍组按照音高进行语气识别的比率也低于儿童控制组。这体现出儿童音高障碍组对于音高的感知整体差于儿童控制组。在声调任务中，对于自然刺激，儿童音高障碍组能够正常判

断声调，与儿童控制组没有显著差异。而在语气音高互换了的不自然刺激中，儿童音高障碍组的声调判断差于儿童控制组。说明语气中音高互换的操作（使音高与声强不一致）明显干扰了儿童音高障碍组对于声调的识别，可以推测儿童音高障碍组对于声强因素比较敏感，而声强本身是语气中的一个线索，所以这个结果提示我们可能儿童尚需依赖语气进行声调判断，即使在判断声调的任务中也加工了语气中的声强线索。另外，从 t 检验可以看到，两组儿童都能够识别大部分刺激类型的语气，但无论是儿童控制组还是音高障碍组，都不能成功识别出带有疑问音高的四声声调，说明这个年龄的儿童在这类带有冲突的刺激上无法顺利进行声调判断。

将成人和儿童的控制组进行对比也发现，儿童控制组的声调识别率显著低于成人控制组，而且只对儿童控制组，当刺激是疑问音高时，四声的正确率显著差于二声，在成人控制组中并没有发现这样的音高和声调间的交互作用。这也说明这一阶段的儿童在声调加工能力上的发展还不够成熟，与成年人的水平有差异，并且还没有掌握可以不受语气干扰地识别出声调的能力。在语气识别上，儿童控制组在八种刺激上和成人控制组一样，都显著高于随机水平，说明九岁的儿童控制组能够进行语气加工，而且和成人控制组一样也是按照音高进行的语气判断。值得一提的是，对两组的方差分析发现疑问语气比陈述语气识别率更低的现象只在成人控制组中出现，而儿童控制组中没有出现，这提示我们九岁的儿童还没有形成陈述语气偏向（statement bias），这种偏向可能是一种帮助进行更有效的语气加工的策略，将陈述语气当作一种默认语气，而疑问语气则标记上需要满足的一些特征，目前看来儿童虽然具备了成熟的语气识别的能力，仍尚未形成这种成人使用的加工策略。

# 第 5 章 总结

## 5.1 研究结论及意义

本研究通过两个实验，探究了与汉语语气和声调加工的一系列问题。包括了（1）在汉语句子的加工过程中，语气和声调是否存在时序层级上的差异，何者是更先被大脑进行处理的变量？符合任务相关假设还是语气优先假设？（2）汉语母语的被试在进行语气判断的时候是否主要依靠音高线索进行识别？包括声强在内的其他线索是否及如何发挥作用？（3）在汉语句子中，语气是否影响声调的识别？声调是否影响语气的识别？（4）音乐障碍的被试群体是如何加工语气与声调的？音乐障碍是否影响语气和声调的加工？

### 5.1.1 语气和声调在时序层级上的差异

研究结果发现，在进行语气加工时，语气是优先诱发出脑电反应的因素，而在进行声调加工时，大脑在非常早期就同时加工了语气和声调因素，这两者存在交互作用，被试对声调刺激的反应也是建立在语气识别的基础上的，总体看来任务相关假设不成立，语气优先假设也没有完全符合，但结果整体还是更倾向于支持语气的优先加工地位。由于语气对于满足人类社会交往需求和进一步生存的重要性[25]，以及它可能具有的更久远的进化根源，它很可能是人类更加敏感和优先加工的一类刺激。未来需要更多研究探讨这一问题，用更多的技术和方法考察语气是否在加工中处于优先地位。

### 5.1.2 语气和声调的成熟早晚差异

对儿童的语气和声调加工的研究发现，九岁的儿童已经具有比较成熟的语气识别能力了，可以成功对不同类型的刺激材料进行语气判断，不会受到声调的干扰。但与此同时，他们的声调加工还发展得不完全，不能够成功识别出具有疑问语气的四声声调。这一结果提示我们，声调很有可能比语气能力成熟得更晚，至少对于九岁的儿童，目前观测到了更早成熟的语气加工能力。语气在发育时间上这一优势或许跟前文提到过的它对于人类社交的重要意义有关系。希望未来有更多早于九岁的相关研究，探察儿童语气加工和声调加工在各个年龄段的具体状况，并进一步探究声调与语气的相对发育过程及可能存在的交互作用。

### 5.1.3 音高、声强、声调对语气识别的影响

实验结果证明，汉语母语被试在进行语气判断时主要依靠音高进行加工，但声强等

其他线索依然有作用，当音高和声强等线索指示同一个语气时识别率最高，当音高和声强等线索指示不同的语气时，大部分情况下被试都倾向于按照音高进行判断。其次，语气和声调互相影响，例如，当语气是疑问语气，而声调是下降的四声时，更加有利于成功识别出疑问语气，但对于声调的识别却加大了难度。无论是音乐能力正常的被试或是音高加工障碍者都是这样的，并且这两类群体都主要以音高进行语气和声调加工的线索。对于音乐能力好或差的被试，语气和声调都是存在相互影响的效应的。

### 5.1.4 音乐障碍对语气和声调加工的影响

虽然语气和声调都以 F0 为主要识别因素，但成人和儿童的实验结果发现，音高能力的障碍会主要影响到语气的识别，对于声调加工的影响不大。音乐障碍的成人在语气加工上的问题与前人的发现一致[64, 69]。关于音乐障碍对儿童影响的研究较少，但相似的语气感知的结果也在人工耳蜗研究中发现[83]。在 See 和同事的实验中，带人工耳蜗的被试（平均年龄为 10.57 岁）在语气识别上的准确率低于正常听力儿童（平均年龄为 7.96 岁）。由于带人工耳蜗的被试在音高感知上也受到了影响，所以这一结果与本研究中音乐能力低下儿童的结果有相似之处，可能语气识别需要精确的音高感知水平，造成了这两个特殊的儿童被试群体较低的识别准确率，也造成了音乐障碍的成人较低的语气识别准确率。而在声调识别（在本研究中即二声和四声的识别）的任务中，对于儿童，虽然有他们不能判断的声调——含有疑问音高的四声刺激，但两组儿童都不能判断出，是一个跟年龄相关的整体问题。音乐障碍的成人和儿童并没有比正常被试在声调识别上有更大的困难，可能是由于声调判断是一个后天逐步习得的能力，先天的音高加工障碍对其没有产生太大影响，因此受到音乐缺陷的影响就没有那么大。



## 5.2 未来研究方向

语气和声调对于人类理解彼此对话的内容和意义都有着非常重要的作用，关于语气和声调的加工机制在未来应有更多的研究。一个重要的探索方向就是语气和声调的加工优先级，例如在普通话除外的其他语言和方言中，比如非声调语言（英语）以及声调语言（粤语）中，是否语气是在早期时间窗就优先被大脑加工的因素。同时，呼吁有更多更前沿的脑科学研究手段和工具来探索这一问题。其次，是否所有群体都对语气敏感也是一个值得探讨的问题。比如，失乐症群体是否在语气和声调任务中也都首先就加工了语气线索，与正常音乐能力的被试有无时间层次上的差异，这都是未来可以研究的方向。

此外，目前本研究在九岁左右（小学三年级）的儿童群体中发现了尚未发展成熟的声调加工能力，那么到底声调加工是在什么时候发育成熟，行为水平不再受到诸如语气

这样因素的干扰，也是非常有意义的研究问题。另外，本研究发现了成人具有、九岁儿童群体中尚未发现的陈述语气偏向（statement bias），那它又是在什么时间发展出的呢？未来应该有更多研究回答这些关键问题，可以着眼于年龄稍大些的儿童，也希望能够有长期追踪研究关注这一发展进程。

# 参考文献

[1] MEYER M, STEINHAUER K, ALTER K, et al. Brain activity varies with modulation of dynamic pitch variance in sentence melody[J]. Brain and language, 2004, 89(2): 277-89.
[2] STEINHAUER K, ALTER K, FRIEDERICI A D. Brain potentials indicate immediate use of prosodic cues in natural speech processing[J]. Nature neuroscience, 1999, 2(2): 191.
[3] NAN Y, FRIEDERICI A D. Differential roles of right temporal cortex and Broca's area in pitch processing: evidence from music and Mandarin[J]. Hum Brain Mapp, 2013, 34(9): 2045-54.
[4] CONNELL B A, HOGAN J T, ROZSYPAL A J. Experimental-Evidence of Interaction between Tone and Intonation in Mandarin Chinese[J]. Journal of Phonetics, 1983, 11(4): 337-51.
[5] YUAN J. Mechanisms of question intonation in Mandarin[M]. // Chinese spoken language processing. City: Springer, 2006: 19-30.
[6] YUAN J. Perception of intonation in Mandarin Chinese[J]. J Acoust Soc Am, 2011, 130(6): 4063-9.
[7] MA J K-Y, CIOCCA V, WHITEHILL T L. The perception of intonation questions and statements in Cantonese[J]. The Journal of the Acoustical Society of America, 2011, 129(2): 1012-23.
[8] REN G-Q, TANG Y-Y, LI X-Q, et al. Pre-attentive processing of mandarin tone and intonation: evidence from event-related potentials[M]. // Functional Brain Mapping and the Endeavor to Understand the Working Brain. City: InTech, 2013.
[9] REN G Q, YANG Y, LI X. Early cortical processing of linguistic pitch patterns as revealed by the mismatch negativity[J]. Neuroscience, 2009, 162(1): 87-95.
[10] LIU M, CHEN Y, SCHILLER N O. Online processing of tone and intonation in Mandarin: Evidence from ERPs[J]. Neuropsychologia, 2016, 91: 307-17.
[11] KUNG C, CHWILLA D J, SCHRIEFERS H. The interaction of lexical tone, intonation and semantic context in on-line spoken word recognition: an ERP study on Cantonese Chinese[J]. Neuropsychologia, 2014, 53: 293-309.
[12] ZATORRE R J, EVANS A C, MEYER E, et al. Lateralization of phonetic and pitch discrimination in speech processing[J]. Science, 1992, 256(5058): 846-9.
[13] NAOI N, WATANABE S, MAEKAWA K, et al. Prosody discrimination by songbirds (Padda oryzivora)[J]. PloS one, 2012, 7(10): e47446.
[14] SPIERINGS M J, TEN CATE C. Zebra finches are sensitive to prosodic features of human speech[J]. P Roy Soc B-Biol Sci, 2014, 281(1787).
[15] ARAKI M, BANDI M M, YAZAKI-SUGIYAMA Y. Mind the gap: Neural coding of species identity in birdsong prosody[J]. Science, 2016, 354(6317): 1282-7.
[16] FERNALD A. Approval and Disapproval - Infant Responsiveness to Vocal Affect in Familiar and Unfamiliar Languages[J]. Child Development, 1993, 64(3): 657-74.
[17] FERNALD A. Human maternal vocalizations to infants as biologically relevant signals: an evolutionary perspective[M]. // BARKOW J H, COSMIDES L, TOOBY J. The adapted mind: Evolutionary psychology and the generation of culture. City: Oxford University Press, USA, 1995: 391.
[18] JUSCZYK P W, CUTLER A, REDANZ N J. Infants Preference for the Predominant Stress Patterns of English Words[J]. Child Development, 1993, 64(3): 675-87.
[19] JUSCZYK P W. How infants begin to extract words From speech[J]. Trends in cognitive sciences, 1999, 3(9): 323-8.
[20] HIRST D, DI CRISTO A. Intonation systems: a survey of twenty languages[M]. City: Cambridge University Press, 1998.

[21] CRUTTENDEN A. Intonation[M]. City: Cambridge University Press, 1997.
[22] BANSE R, SCHERER K R. Acoustic profiles in vocal emotion expression[J]. Journal of Personality and Social Psychology, 1996, 70(3): 614-36.
[23] LADD D R. Intonational Phonology[M]. City: Cambridge University Press, 1996.
[24] THOMPSON W F, BALKWILL L L. Decoding speech prosody in five languages[J]. Semiotica, 2006, 158(1-4): 407-24.
[25] BRANCUCCI A, LUCCI G, MAZZATENTA A, et al. Asymmetries of the human social brain in the visual, auditory and chemical modalities[J]. Philos T R Soc B, 2009, 364(1519): 895-914.
[26] XU Y S, GANDOUR J T, FRANCIS A L. Effects of language experience and stimulus complexity on the categorical perception of pitch direction[J]. Journal of the Acoustical Society of America, 2006, 120(2): 1063-74.
[27] HALLE P A, CHANG Y C, BEST C T. Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners[J]. Journal of Phonetics, 2004, 32(3): 395-421.
[28] LIU C, RODRIGUEZ A. Categorical perception of intonation contrasts: Effects of listeners' language background[J]. Journal of the Acoustical Society of America, 2012, 131(6): El427-El33.
[29] HO A T. Intonation variation in a Mandarin sentence for three expressions: interrogative, exclamatory and declarative[J]. Phonetica, 1977, 34(6): 446-57.
[30] LIU H M, TSAO F M, KUHL P K. Acoustic analysis of lexical tone in Mandarin infant-directed speech[J]. Developmental psychology, 2007, 43(4): 912-7.
[31] MA J K, CIOCCA V, WHITEHILL T L. Effect of intonation on cantonese lexical tones[J]. J Acoust Soc Am, 2006, 120(6): 3978-87.
[32] NETTL B. An ethnomusicologist contemplates universals in musical sound and musical culture[J]. Origins of Music, 2000: 463-72.
[33] PATEL A D. Language, music, syntax and the brain[J]. Nature neuroscience, 2003, 6(7): 674-81.
[34] PATEL A D. Music, Language, and the Brain[J]. 2010.
[35] BUTZLAFF R. Can music be used to teach reading?[J]. J Aesthet Educ, 2000, 34(3-4): 167-78.
[36] CHAN A S, HO Y C, CHEUNG M C. Music training improves verbal memory[J]. Nature, 1998, 396(6707): 128.
[37] HO Y-C, CHEUNG M-C, CHAN A S. Music training improves verbal but not visual memory: Cross-sectional and longitudinal explorations in children[J]. Neuropsychology, 2003, 17(3): 439-50.
[38] MORENO S, BIALYSTOK E, BARAC R, et al. Short-term music training enhances verbal intelligence and executive function[J]. Psychological science, 2011, 22(11): 1425-33.
[39] GARDINER M F, FOX A, KNOWLES F, et al. Learning improved by arts training[J]. Nature, 1996, 381(6580): 284-.
[40] COSTA-GIOMI E. The effects of three years of piano instruction on children's cognitive development[J]. J Res Music Educ, 1999, 47(3): 198-212.
[41] HETLAND L. Learning to make music enhances spatial reasoning[J]. J Aesthet Educ, 2000, 34(3-4): 179-238.
[42] RAUSCHER F H, ZUPAN M A. Classroom keyboard instruction improves kindergarten children's spatial-temporal performance: A field experiment[J]. Early Child Res Q, 2000, 15(2): 215-28.
[43] TAI T-C. The effect of violin, keyboard, and singing instruction on the spatial ability and music aptitude of young children[J]. 2010.
[44] ZAFRANAS N. Piano keyboard training and the spatial–temporal development of young children attending kindergarten classes in Greece[J]. Early Child Development and Care, 2004, 174(2): 199-211.
[45] KALMUS H, FRY D. On tune deafness (dysmelodia): frequency, development, genetics and musical background[J]. Annals of human genetics, 1980, 43(4): 369-82.
[46] PERETZ I. Brain specialization for music. New evidence from congenital amusia[J]. Annals of the New York Academy of Sciences, 2001, 930: 153-65.

[47] NAN Y, SUN Y, PERETZ I. Congenital amusia in speakers of a tone language: association with lexical tone agnosia[J]. Brain, 2010: awq178.
[48] DRAYNA D, MANICHAIKUL A, DE LANGE M, et al. Genetic correlates of musical pitch recognition in humans[J]. Science, 2001, 291(5510): 1969-72.
[49] PERETZ I, CUMMINGS S, DUBE M P. The genetics of congenital amusia (tone deafness): A family-aggregation study[J]. Am J Hum Genet, 2007, 81(3): 582-8.
[50] HYDE K L, PERETZ I. Brains that are out of tune but in time[J]. Psychological science, 2004, 15(5): 356-60.
[51] AYOTTE J, PERETZ I, HYDE K. Congenital amusia: a group study of adults afflicted with a music-specific disorder[J]. Brain, 2002, 125(2): 238-51.
[52] PERETZ I, CHAMPOD A, HYDE K. Varieties of musical disorders. The Montreal Battery of Evaluation of Amusia[J]. Annals of the New York Academy of Sciences, 2003, 999: 58-75.
[53] PATEL A D, WONG M, FOXTON J, et al. Speech intonation perception deficits in musical tone deafness (congenital amusia)[J]. Music Perception: An Interdisciplinary Journal, 2008, 25(4): 357-68.
[54] PATEL A D, GIBSON E, RATNER J, et al. Processing Syntactic Relations in Language and Music: An Event-Related Potential Study[J]. J Cognitive Neurosci, 1998, 10(6): 717-33.
[55] PATEL A D. An Empirical Method for Comparing Pitch Patterns in Spoken and Musical Melodies: A Comment on JGS Pearl's "Eavesdropping with a Master: Leos Janáćek and the Music of Speech"[J]. Empirical Musicology Review, 2006, 1(3).
[56] PLANTINGA J, TRAINOR L J. Memory for melody: Infants use a relative pitch code[J]. Cognition, 2005, 98(1): 1-11.
[57] BIDELMAN G M, GANDOUR J T, KRISHNAN A. Cross-domain Effects of Music and Language Experience on the Representation of Pitch in the Human Auditory Brainstem[J]. J Cognitive Neurosci, 2011, 23(2): 425-34.
[58] BIDELMAN G M, HUTKA S, MORENO S. Tone Language Speakers and Musicians Share Enhanced Perceptual and Cognitive Abilities for Musical Pitch: Evidence for Bidirectionality between the Domains of Language and Music[J]. PloS one, 2013, 8(4).
[59] WONG P C, SKOE E, RUSSO N M, et al. Musical experience shapes human brainstem encoding of linguistic pitch patterns[J]. Nature neuroscience, 2007, 10(4): 420-2.
[60] JIANG C, HAMM J P, LIM V K, et al. Amusia results in abnormal brain activity following inappropriate intonation during speech comprehension[J]. PloS one, 2012, 7(7): e41411.
[61] JONES J L, ZALEWSKI C, BREWER C, et al. Widespread Auditory Deficits in Tune Deafness[J]. Ear Hearing, 2009, 30(1): 63-72.
[62] LIU F, JIANG C, THOMPSON W F, et al. The mechanism of speech processing in congenital amusia: Evidence from Mandarin speakers[J]. PloS one, 2012, 7(2): e30374.
[63] TILLMANN B, BURNHAM D, NGUYEN S, et al. Congenital amusia (or tone-deafness) interferes with pitch processing in tone languages[J]. Frontiers in psychology, 2011, 2.
[64] LIU F, PATEL A D, FOURCIN A, et al. Intonation processing in congenital amusia: discrimination, identification and imitation[J]. Brain, 2010, 133(Pt 6): 1682-93.
[65] LIU F, JIANG C, WANG B, et al. A music perception disorder (congenital amusia) influences speech comprehension[J]. Neuropsychologia, 2015, 66: 111-8.
[66] PATEL A D, FOXTON J M, GRIFFITHS T D. Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech[J]. Brain and cognition, 2005, 59(3): 310-3.
[67] NGUYEN S, TILLMANN B, GOSSELIN N, et al. Tonal language processing in congenital amusia[J]. Annals of the New York Academy of Sciences, 2009, 1169: 490-3.
[68] HUTCHINS S, GOSSELIN N, PERETZ I. Identification of changes along a continuum of speech intonation is

impaired in congenital amusia[J]. Frontiers in psychology, 2010, 1: 236.
[69] JIANG C, HAMM J P, LIM V K, et al. Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese[J]. Neuropsychologia, 2010, 48(9): 2630-9.
[70] LIANG J, VAN HEUVEN V J. Chinese tone and intonation perceived by L1 and L2 listeners[J]. Tones and tunes, 2007, 2: 27-61.
[71] PENG S C, LU N, CHATTERJEE M. Effects of cooperating and conflicting cues on speech intonation recognition by cochlear implant users and normal hearing listeners[J]. Audiol Neurootol, 2009, 14(5): 327-37.
[72] OLDFIELD R C. The assessment and analysis of handedness: the Edinburgh inventory[J]. Neuropsychologia, 1971, 9(1): 97-113.
[73] BOERSMA P, WEENINK D. Praat [Computer Software], Version 4.3[J]. Amsterdam, Netherlands: Institute of Phonetic Sciences, University of Amsterdam, 2004.
[74] SEMLITSCH H V, ANDERER P, SCHUSTER P, et al. A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP[J]. Psychophysiology, 1986, 23(6): 695-703.
[75] KOLK H H, CHWILLA D J, VAN HERTEN M, et al. Structure and limited capacity in verbal working memory: A study with event-related potentials[J]. Brain and language, 2003, 85(1): 1-36.
[76] KOLK H, CHWILLA D. Late positivities in unusual situations[J]. Brain and language, 2007, 100(3): 257-61.
[77] LI X, GANDOUR J, TALAVAGE T, et al. Selective attention to lexical tones recruits left dorsal frontoparietal network[J]. Neuroreport, 2003, 14(17): 2263-6.
[78] GANDOUR J, DZEMIDZIC M, WONG D, et al. Temporal integration of speech prosody is shaped by language experience: An fMRI study[J]. Brain and language, 2003, 84(3): 318-36.
[79] IACOBONI M, LIEBERMAN M D, KNOWLTON B J, et al. Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline[J]. NeuroImage, 2004, 21(3): 1167-73.
[80] GONG Y. Wechsler adult intelligence scale-revised in China Version[J]. Hunan Medical College, Changsha, Hunan/China, 1992.
[81] PERETZ I, GOSSELIN N, NAN Y, et al. A novel tool for evaluating children's musical abilities across age and culture[J]. Frontiers in systems neuroscience, 2013, 7: 30.
[82] RAVEN J C. Standardization of Progressive Matrices, 1938[J]. Brit J Med Psychol, 1941, 19: 137-50.
[83] SEE R L, DRISCOLL V D, GFELLER K, et al. Speech Intonation and Melodic Contour Recognition in Children With Cochlear Implants and With Normal Hearing[J]. Otol Neurotol, 2013, 34(3): 490-8.